# Identification of regulators and effectors of single-cell $C_4$ morphology by comparative transcriptomics of dark-grown cotyledons of the single-cell $C_4$ plant *Bienertia sinuspersici* and the related $C_3$ plant *Suaeda heterophylla*

Von der Naturwissenschaftlichen Fakultät der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades
Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation
von
Diplom-Biologin Lisa Hagenau

2017

Referent:             Dr. Sascha Offermann
Korreferent:          Prof. Dr. Helge Küster
Tag der Promotion:    01.06.2017

# Abstract

*Bienertia sinuspersici* is one of four plants currently known to perform $C_4$ photosynthesis within individual cells (single-cell $C_4$ ($SCC_4$)). The required spatial separation for the primary and secondary carbon fixation is accomplished by developing two functionally distinct subcellular domains, the central and the peripheral compartment, each containing one type of dimorphic chloroplast. Three developmental stages have been described previously: young cells do not show subcellular compartimentation, while cells in the intermediate stage show aggregation of chloroplasts into a precentral compartment. Mature cells have distinct subcellular compartments separated by a large vacuole and connected by cytoplasmic strands. The subcellular organelle arrangement is maintained by actin and microtubules and is only disturbed in prolonged low light conditions. However, the mechanism and the transcriptional regulation behind the development of this unique cell morphology are still unknown. The aim of this thesis was to identify factors involved in the development of the two subcellular compartments in *B. sinuspersici*. A comparative RNA sequencing experiment spanning all stages of $SCC_4$ development was designed to identify genes involved in $SCC_4$ morphology and to explore cotyledon development in dark-grown seedlings of *B. sinuspersici* and the related $C_3$ species *Suaeda heterophylla*. It is shown that the spatial separation of the dimorphic compartments develops in a light-independent manner in dark-grown *B. sinuspersici* cotyledons. Several read mapping strategies were tested and evaluated for their suitability in comparing gene expression between *B. sinuspersici* and *S. heterophylla*. A protein-based read mapping to the transcriptome of *B. vulgaris* is shown to be the best tool for a cross-species comparison. Multiple genes show changes in gene expression that correspond to the observed morphological changes in cell development. CML24, IQD2 and MASP1 have previously been shown to play a role in microtubule organisation and were identified as potential factors in SCC4 development here. In addition, GFP-expression studies of a truncated chloroplast movement gene, CHUP1, disturbs the positioning of peripheral chloroplasts. Finally, multiple transcription factors are highly abundant in *B. sinuspersici*, but not *S. heterophylla*. These are identified as potential regulators of $SCC_4$ development. The identified genes can serve as direct targets for functional analysis in the future.

**Keywords:** single-cell C4 photosynthesis, RNA sequencing, comparative transcriptomics

# Zusammenfassung

*Bienertia sinuspersici* ist eine von vier Pflanzen, die $C_4$ Photosynthese innerhalb einzelner Zellen durchführen können (single-cell $C_4$ ($SCC_4$)). Die erforderliche räumliche Trennung für die primäre und sekundäre Kohlenstofffixierung wird durch die Entwicklung von zwei funktionell getrennten subzellulären Domänen, dem zentralen und dem peripheren Kompartiment, erreicht, die jeweils einen dimorphen Chloroplastentyp enthalten. Es sind drei Zellentwicklungsstadien beschrieben: Junge Zellen zeigen keine subzelluläre Kompartimentierung, während in älteren Zellen die Aggregation von Chloroplasten in ein prä-zentrales Kompartiment zu beobachten ist. Vollentwickelte Zellen besitzen zwei subzelluläre Kompartimente, die durch eine große Vakuole getrennt und durch zytoplasmatische Stränge verbunden sind. Die subzelluläre Organellenanordnung wird durch Aktin und Mikrotubuli aufrechterhalten und nur durch längere Schwachlichtverhältnisse gestört. Der Mechanismus und die Transkriptionsregulation der Entwicklung dieser einzigartigen Zellmorphologie sind allerdings noch unbekannt. Ziel dieser Arbeit war es, Faktoren zu identifizieren, die an der Entwicklung der beiden subzellulären Kompartimente in *B. sinuspersici* beteiligt sind. Ein comparative RNA sequencing Experiment, das alle Stadien der $SCC_4$-Entwicklung umfasst, wurde designt, um Gene zu identifizieren, die an der Bildung der $SCC_4$-Morphologie beteiligt sind, und die Entwicklung von Kotyledonen in dunkelgewachsenen Keimlingen von *B. sinuspersici* und der verwandten $C_3$-Pflanze *Suaeda heterophylla* zu erforschen. Es wird gezeigt, daß in dunkelgewachsenen *B. sinuspersici* Kotyledonen die räumliche Trennung der dimorphen Kompartimente lichtunabhängig erfolgt. Mehrere Read Mapping-Strategien wurden auf ihre Eignung geprüft, die Genexpression zwischen *B. sinuspersici* und *S. heterophylla* zu vergleichen. Ein Protein-basiertes Read-Mapping auf das Transkriptom von *B. vulgaris* erwies sich hierfür als das beste Werkzeug. Mehrere Gene zeigen Expressionsänderungen, die den beobachteten morphologischen Veränderungen in der Zellentwicklung entsprechen. Den Genen CML24, IQD2 und MASP1, die hier als potentielle Faktoren in der SCC4-Entwicklung identifiziert werden, wurde zuvor eine Rolle in der Mikrotubuli-Organisation nachgewiesen. Weiterhin wird gezeigt, dass die Expression eines trunkierten CHUP1-GFP-Fusionsproteins die Positionierung von peripheren Chloroplasten stört. Zudem sind mehrere Transkriptionsfaktoren in *B. sinuspersici* höher exprimiert als in *S. heterophylla*. Diese werden als potenzielle Regulatoren der $SCC_4$ Entwicklung identifiziert. Die identifizierten Gene können in Zukunft als direkte Ziele für die funktionale Analyse dienen.

**Schlagworte:** single-cell C4 Photosynthese, RNA-Sequenzierung, vergleichende Transkriptomanalyse

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| ABA | abscisic acid |
| ACT | actin |
| BASS | bile acid:sodium symporter |
| Bienertia | *Bienertia sinuspersici* |
| BS | bundle sheath |
| CA | carbonic anhydrase |
| CBB | Calvin-Benson-Bassham |
| CC | central compartment |
| CCCP | central compartment chloroplast |
| CHUP | chloroplast unusual positioning |
| CML | calmodulin-like |
| $CO_2$ | carbon dioxide |
| CV | coefficient of variation |
| DNA | deoxyribonucleic acid |
| GFP | green fluorescent protein |
| KAC | kinesin-like protein for actin-based chloroplast movement |
| M | mesophyll |
| MASP | microtubule-associated stress protein |
| NAD-ME | NAD-malic enzyme |
| NC | normalized counts |
| OAA | oxalacetate |
| ORF | open reading frame |
| PC | peripheral compartment |
| PCA | principal component analysis |
| PCP | peripheral compartment chloroplast |
| PCR | polymerase chain reaction |
| PEPC | phosphoenolpyruvate carboxylase |

| | |
|---|---|
| PPDK | pyruvate phosphate dikinase |
| qPCR | quantitative real-time PCR |
| RNA | ribonucleic acid |
| RPKM | reads per kilobase per million mapped reads |
| RuBisCO | ribulose-1,5-biphosphate carboxylase oxygenase |
| $SCC_4$ | single cell $C_4$ |
| Seq | sequencing |
| Suaeda | *Suaeda heterophylla* |
| TC | total counts |
| TP | time point |

# 1 Introduction

The production of dephlogisticated air from leaves is not owing to the warmth of the sun, but chiefly, if not only, to the light. No dephlogisticated air is obtained in a warm room, if the sun does not shine upon the jar containing the leaves.

*(Jan IngenHousz, 1779)*

## 1.1 Photosynthesis and the dual function of RuBisCO

Jan IngenHousz was a pioneer of photosynthesis research and contributed to discovering the arguably most important biochemical process in 1779. Since then, countless scientists have investigated the underlying biochemistry, physiology and regulation of photosynthesis. The signature enzyme of photosynthesis is ribulose-1,5-biphosphate-carboxylase/oxygenase (RuBisCO), which catalyzes the reaction of carbon dioxide ($CO_2$) and ribulose-1,5-biphosphate to 3-Phosphoglyeric acid (3-PGA), the first step of the Calvin-Benson-Bassham (CBB) cycle. While RuBisCO has a high affinity to $CO_2$, it can also use $O_2$ as a substrate. The resulting phosphoglycolate is toxic and further metabolized in mitochondria and peroxisomes to glycerate, which can then re-enter the CBB-cycle as 3-PGA. Briefly, photorespiration leads to a net loss of $CO_2$ and consumption of $O_2$ (Ogren, 1984). Under the atmospheric conditions that prevailed when photosynthesis first developed, this trait was of little importance, as $CO_2$ concentrations were much higher than $O_2$ concentrations (Ehleringer et al. 1991). In current atmospheric conditions, especially in warmer temperatures, photorespiration can inhibit photosynthesis by over 30 % (Sage 2004).

## 1.2 C$_4$ photosynthesis

Several plants have developed mechanisms that reduce the likelihood of photores-
piration taking place by increasing the local concentration of CO$_2$ around RuBisCO.
One of these adaptations is C$_4$ photosynthesis, which is usually reflected in the leaf
anatomy of the plant: The so-called Kranz anatomy refers to a specific arrangement
of veins, bundle sheath (BS) cells, and mesophyll (M) cells within the leaf. Veins are
surrounded by large BS cells, which are in turn surrounded by M cells, and are thus
separated from the next veins by at most two BS and 2 M cells. In C$_3$ plants, veins
are spaced much wider and are separated by many M cells (Dengler et al. 1994; Ogle
2003). M and BS cells work together to perform C$_4$ photosynthesis. In M cells, CO$_2$ is
fixed preliminary to phosphoenolpyruvate (PEP), a C$_3$-molecule, to form oxaloacetic
acid (OAA), a C$_4$-molecule, hence the name C$_4$ photosynthesis. OAA is transported
to the BS cells and decarboxylated in the chloroplasts containing RuBisCO, which
catalyzes the second and final CO$_2$ fixation step in the CBB cycle. There are three
main variants of the C$_4$ pathway, each is named for the enzyme that decarboxylates
the C$_4$ acid in BS cells (Hatch 1987). The NADP-ME pathway occurs in maize (*Zea
mays*) and *Sorghum bicolor*, while *Amaranthus hypochondriacus* and *Eleusine coracana* are
NAD-ME-type C$_4$ species. The third C$_4$-type, PEPCK, occurs the least often and the
question whether it really is an independent C$_4$ subtype has not yet been resolved
(Sage 2004; Bräutigam et al. 2014). C$_4$ photosynthesis evolved on at least 61 sepa-
rate occasions and the habitat of C$_4$ plants is often, but not necessarily, arid or warm
climates (Sage 2016).

## 1.3 Bienertia - a single-cell C$_4$ photosynthesis performing plant

Roughly a decade ago, the phenomenon of single-cell C$_4$ (SCC$_4$) photosynthesis was
discovered. *Suaeda aralocaspica* and *Bienertia cycloptera* were found to have C$_4$-like
$\delta^{13}$C values without the presence of "Kranz anatomy", which until then had been
thought to be essential for C$_4$ photosynthesis (Freitag et al. 2000; Voznesenskaya et
al. 2001; Freitag et al. 2002; Voznesenskaya et al. 2002). Since then, SCC$_4$ photosyn-
thesis has been discovered in two additional species, *Bienertia sinuspersici* (Bienertia)

peripheral chloroplast

central compartment

vacuole

peroxisomes

mitochondria

distal compartment

proximal compartment

*Bienertia sinuspersici*          *Suaeda aralocaspica*
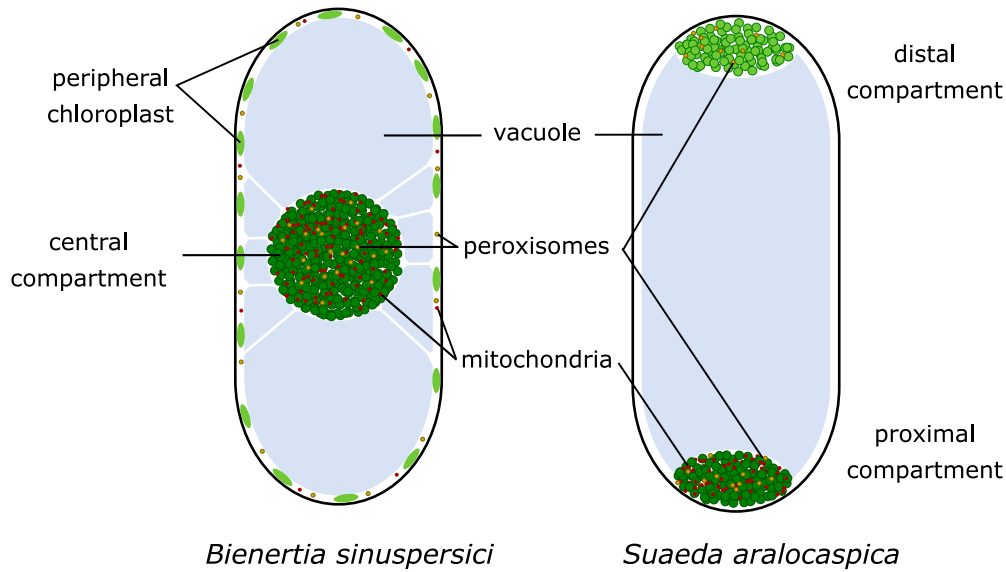
Figure 1.1: Organization of dimorphic compartments in chlorenchyma cells of the $SCC_4$ species Bienertia (left) and Suaeda (right). In Bienertia, the central compartment (CC) is functionally equivalent to Kranz $C_4$ BS cells and contains central compartment chloroplasts (dark green), the majority of mitochondria (red) and peroxisomes (yellow). The peripheral compartment (PC) contain peripheral chloroplasts (light green) and some scattered mitochondria and peroxisomes. Both compartments are spatially separated by the vacuole. In Suaeda, the proximal compartment is functionally synonymous with the CC in Bienertia and contains proximal chloroplasts (dark green), all mitochondria and peroxisomes. The distal compartment is the equivalent of the PC in Bienertia and contains distal chloroplasts (light green) and some peroxisomes.

and *Bienertia kavirense* (Akhani et al. 2005; Akhani et al. 2012). In the $SCC_4$ species, the functions of M and BS cells are combined into one cell with two distinct compartments each with a specialized chloroplast type. In *Suaeda aralocaspica*, the two compartments are localized at opposite ends of the cell (Voznesenskaya et al. 2001), while in Bienertia, the spatial separation is between cell center and periphery (Freitag et al. 2002) (Figure 1.1). The central compartment chloroplasts (CCCP), functionally synonymous to the Kranz-$C_4$ BS chloroplasts, form a ball-like structure in the center of the cell, while the peripheral chloroplasts (PCP) are distributed evenly along the cell wall (Voznesenskaya et al. 2002). Their main function is the generation of the primary $CO_2$-acceptor PEP (Voznesenskaya et al. 2002; Offermann et al. 2011). Both mitochondria and peroxisomes are predominantly located in the central compartment (CC) with only a few scattered around the periphery (Voznesenskaya et al. 2005; Chuong et al. 2006; Park et al. 2009).

3

### 1.3.1 C$_4$-type

The physiological and biochemical side of SCC$_4$ photosynthesis has been studied extensively. The biochemistry of the C$_4$ pathway in Bienertia differs very little from that of the Kranz C$_4$ species. Bienertia belongs to the NAD-ME subtype of C$_4$ plants (Offermann et al. 2011). CO$_2$ enters the cell by diffusion and is fixed to PEP by the PEP-carboxylase (PEPC) to yield OAA. OAA is transaminated to aspartate by the cytosolic aspartate aminotransferase (cAsp-AT) and diffuses into the mitochondria of the CC, where it is de-aminated by the mitochondrial aspartate aminotransferase (mAsp-AT). The resulting OAA is reduced to malate by the NAD-malate dehydrogenase (NAD-MDH), which is then de-carboxylated by the NAD-malic enzyme (NAD-ME). The freed CO$_2$ diffuses to the CCCPs where it is fixed by RuBisCO in the C$_3$ cycle. The other product of the decarboxylation of malate is pyruvate, which is transported out of the mitochondria into the cytosol. It is transaminated to alanine, transported to the peripheral compartment (PC) and de-aminated to form pyruvate. It is imported into the PCPs as substrate for the pyruvate phosphate dikinase (PPDK), which regenerates the primary CO$_2$ acceptor PEP (Offermann et al. 2011; Offermann et al. 2015). The existence of a triose-phosphate shuttle between the two chloroplast types has been suggested in order to maintain the energy balance in the cell (Offermann et al. 2011). In summary, the CC is functionally equivalent to BS cells in C$_4$ plants, where CO$_2$ is fixed via RuBisCO. The main function of the PCP is regeneration of the primary CO$_2$ acceptor PEP. Accordingly, several studies have shown that RuBisCO and other CBB genes accumulate preferentially in the CCCPs, while PPDK is selectively accumulating in the PCPs (Voznesenskaya et al. 2002; Offermann et al. 2011). The cell size and the resulting spatial separation of the dimorphic chloroplasts is essential for a functional C$_4$ photosynthetic pathway by reducing CO$_2$ leakage (Jurić et al. 2017). The efficiency of SCC$_4$ photosynthesis is considered to be equal to that of Kranz C$_4$ species (King et al. 2012).

### 1.3.2 SCC$_4$ development

The development of the cell morphology has been well described (Voznesenskaya et al. 2005; Park et al. 2009; Koteyeva et al. 2016). The youngest cells at the base of the leaf are small and contain a large nucleus, several small vacuoles and a number of randomly distributed undifferentiated chloroplasts, mitochondria and peroxisomes.

4
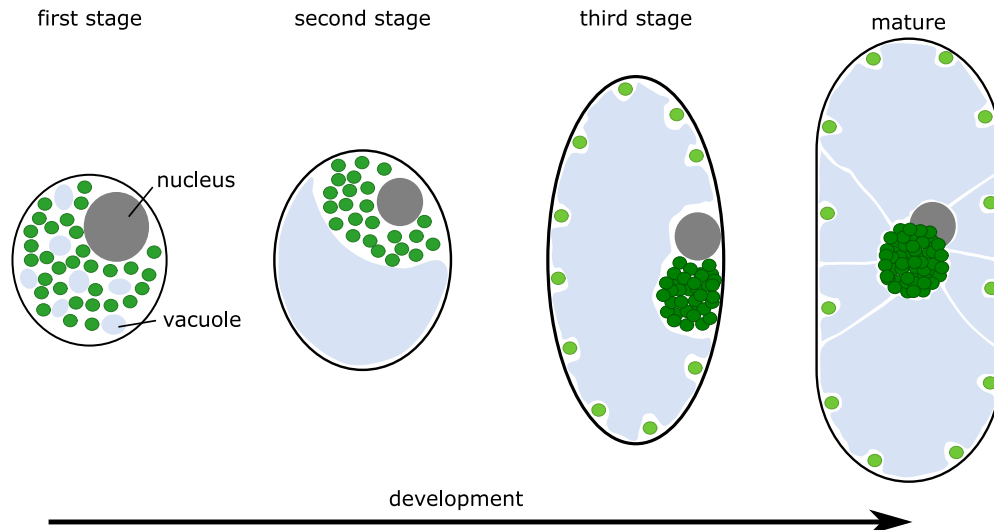
Figure 1.2: Schematic of the development of Bienertia chlorenchyma cells. In the first developmental stage, the nucleus (grey) is large, and several smaller vacuoles (light blue) and chloroplasts (green) are randomly scattered around the cell. In the second stage, the smaller vacuoles have fused into a larger one and chloroplasts aggregate near the nucleus to form a pre-CC. In the third stage, the cell has elongated and the dimorphic compartments are clearly visible and spatially separated. At this stage, differential protein accumulation is measurable (indicated by shades of green). Adopted from Park, 2009.

In the intermediate stage, larger and slightly elongated cells are formed, in which chloroplasts and mitochondria have aggregated to form a pre-CC next to the nucleus. Some chloroplasts are distributed around the cell periphery. The several small vacuoles have fused into one large vacuole, which has expanded. At this point, $SCC_4$ morphology is already clearly visible. Mature cells are large, more elongated and show a clear demarcation of the CC and the cell periphery. The large vacuole fills almost the entire cell, leaving only cytoplasmic strands to connect both cytosolic compartments (Fig 1.2).

### 1.3.3 Differentiation of the chloroplasts

The two chloroplast types differentiate during development and acquire different sets of proteins as well as distinct morphologies. In young leaves, there is one undifferentiated chloroplast type that is randomly distributed within the cell. It contains small amounts of RuBisCO, but no $C_4$ gene transcripts. In intermediate leaves, the two chloroplast types are still indistinguishable in shape and size, although spatial

separation has occured. There are conflicting results about the biochemical differentiation. An early study found a lower amount of RuBisCO in the PCPs (Voznesenskaya et al. 2005), while a more recent study showed RuBisCO large subunit mRNA and protein in both chloroplast types (Koteyeva et al. 2016). In mature cells, the two chloroplast types have accumulated specific sets of protein and are distinctly structured. The CCCPs remain small and contain grana stacks, RuBisCO and the other CBB cycle genes. The PCPs are larger in size, adopting a donut-like shape, and contain little grana. Differentially accumulating proteins are PPDK, bile acid:sodium symporter (BASS) and triose phosphate isomerase (TPI).

### 1.3.4 Suggested mechanisms for differential protein accumulation

Three hypothesis were suggested that could explain the mechanism behind differential protein accumulation: protein or mRNA targeting to specific chloroplast type, and chloroplast-specific protein degradation (Offermann et al. 2011). For selective import into the chloroplasts, both chloroplast types would need specific receptors interacting with the targeted precursor protein and the chloroplast import machinery. There is now some evidence that this is the likely mechanism, as transit peptides of proteins accumulating in PCPs (PPDK, BASS, TPI) contain a motif that ensures selective import into PCPs by blocking import into CCCPs (Wimmer et al. 2017). Motifs responsible for the selective import into the CCCPs have not yet been identfied. The mRNA targeting hypothesis involves selective transport of mRNA to the targeted chloroplast type, where translation and import takes place. In this case, selective import is not controlled by the chloroplast. While an mRNA targeting mechanism has been described in plants for mitochondria, there is currently no evidence for mRNA targeting in $SCC_4$. The third hypothesis suggests that protein import into both chloroplast types happens indiscriminately, followed by selective degradation of proteins. It has not been tested in Bienertia and would require specific proteases, which were not identified in a study comparing the proteome of both chloroplast types (Offermann et al. 2015).

### 1.3.5 Chloroplast positioning in Bienertia

In contrast to most other species, the Bienertia chloroplasts have very limited light-dependent chloroplast movement. The PCPs are firmly anchored in place, while the CC moves toward the light source only after prolonged low light exposure (Lara et al. 2008). After a prolonged dark treatment, the CC moves from the center to one side of the cell. The cytoskeleton is involved in maintaining the position of the dimorphic chloroplasts. The PCPs were shown to be surrounded by a ring of actin strands. Actin filaments are associated with chloroplasts from early developmental stages (Park et al. 2009). It is thus likely that chloroplast movement is actin-dependent. The CC is contained within a cage of microtubules in the center of the cell with the nucleus nearby (Chuong et al. 2006; Park et al. 2009). Previous studies have shown that both microtubules and actin are necessary for stabilizing the CC by disrupting the cytoskeleton with cytochalasin D and oryzalin treatment (Chuong et al. 2006; Park et al. 2009). The structure of microtubules changes throughout cell development. In young cells, microtubules show a crossing pattern. In older cells, the organization of microtubules allows for polarized cell growth by exhibiting a pattern perpendicular to the axis of growth (Park et al. 2009). Recent research on the leaf development of Bienertia suggests that the spatial separation of the two chloroplast types occurs before biochemical and structural differentiation, similar to the BS cell chloroplasts in Kranz $C_4$ species from the same subfamily Suaedoideae (Koteyeva et al. 2016). It is unclear whether the structural and biochemical differences in the dimorphic chloroplasts are influenced by chloroplast position.

## 1.4 Chloroplast positioning in Kranz-$C_4$ plants

In most plants, chloroplasts are distributed evenly around the cell under moderate light conditions. An even chloroplast distribution is also observed in the M cells of $C_4$ plants. However, in the BS cells of $C_4$ plants, chloroplasts localize to a centripetal or centrifugal position. In monocots, the BS chloroplast position is indicative of the $C_4$-subtype, while in nearly all eudicot $C_4$ species, BS cell chloroplasts adopt a centripetal position, regardless of subtype (Carolin et al. 1978; Muhaidat et al. 2007). It is thought that both localizations have an impact on the $C_4$ cycle. The centripetally positioned chloroplasts might prevent $CO_2$ leakage to the M, while the centrifugal

position of chloroplasts is theorized to enhance metabolite exchange between M and BS cells (Hattersley et al. 1981). While $C_4$ M chloroplasts can change their positions in response to light or other stresses, BS chloroplasts maintain their position in changing light conditions (Yamada et al. 2009; Maai et al. 2011). The mechanism behind initiating and maintaining chloroplast localization in these cells is unknown. First steps towards solving the question have been made with maize mutants with an altered SCARECROW gene, which have abnormally developed BS chloroplasts that are not centrifugally located (Slewinski et al. 2012). However, SCARECROW is a major regulator of vascular and BS development and is thus likely to only indirectly influence chloroplast position.

## 1.5 Chloroplast movement

### 1.5.1 Light-dependent movement

In all plant species, chloroplast movement is influenced by light. Chloroplasts move to the side wall of the cell in response to high light (avoidance response) and towards the cell surface in low light conditions (accumulation response) (Zurzycki 1955). It is thought that re-orientation of the chloroplasts optimizes photosynthetic output and protects the chloroplast from photodamage in case of high light (Zurzycki 1955; Kasahara et al. 2002; Tholen et al. 2008). Light-dependent chloroplast movement was shown to be dependent on short actin strands (Kadota et al. 2009) and regulated by phototropins (Sakai et al. 2001; Kagawa et al. 2001; Kong et al. 2013). Several components linking actin and chloroplast movement have been identified. Chloroplast unusual positioning 1 (CHUP1) is essential for chloroplast positioning and movement. It localizes to the chloroplast outer membrane and has been shown to interact with actin and profilin *in vitro* (Oikawa et al. 2003; Schmidt von Braun et al. 2008). A coiled-coil region in CHUP1 presumably anchors chloroplasts to the plasma membrane (Oikawa et al. 2008). Kinesin-Like Protein for Actin-Based Chloroplast Movement1 (KAC1) and KAC2 were also found to play a role in chloroplast movement and positioning. Like CHUP1, KAC1 and KAC2 can bind actin (Suetsugu et al. 2010). Identification of CHUP1 and KAC1 homologs in fern and moss suggest that its function is evolutionary conserved (Suetsugu et al. 2012; Oikawa et al. 2008). In Bienertia, the PCPs are immobile even during strong light exposure. The CC can move towards the light

source during prolonged low light exposure (Lara et al. 2008). The movement is slower than the chloroplast accumulation response in other species (Königer et al. 2012). This is not surprising, given that an entire cell compartment consisting of approximately 350 chloroplasts and mitochondria has to move, instead of one chloroplast. As the CC is held in place by the cytoskeleton, this would require a massive rearrangement of the microtubules and actin strands and consequently a lot of energy. It is possible that short-term fluctuations in light intensity are not enough to trigger an accumulation response or the movement is simply to slow and thus stays unnoticed.

## 1.5.2 Light-independent movement

In contrast to light-induced chloroplast movement responses, light-independent chloroplast movement has rarely been investigated. Several species show distinct chloroplast positioning in the dark. In Arabidopsis M cells, chloroplasts are positioned around the sides of the cell that are bordering other M cells, but not the epidermis. It was suggested that this chloroplast arrangement is non-random and induced by environmental conditions (Trojan et al. 1996). In guard cells of Arabidopsis, dark conditions were shown to induce distinct chloroplast positioning as well (Königer et al. 2010). In the $C_4$ species maize and *A. hypochondriacus*, BS chloroplast localization to the centripetal or centrifugal side of the BS cell occurs in a light-independent manner (Wang et al. 1993; Taniguchi et al. 2003), while in the $C_4$ species *Echinochloa utilis*, *S. bicolor* and *Eriachne aristidea*, BS chloroplasts fail to localize to the centripetal or centrifugal side of the BS cell during development in dark conditions (Maai et al. 2011; Taniguchi et al. 2003). It is unclear, whether there is a common mechanism behind these disparate observations. It is possible that some components of the light-dependent mechanism are also involved in light-independent chloroplast movement. Clearly, more research is needed to identify the underlying signaling pathways and effectors of light-independent chloroplast movement.

## 1.6 Cotyledons

Cotyledons are the first photosynthesizing organ of a plant. In contrast to true leaves, which develop from the shoot apical meristem, cotyledons are formed during embryogenesis. Their principal function is to ensure seedling survival until true leaves have developed (Armstrong et al. 1993). Cotyledons are popular models for leaf morphogenesis, as altered expression of leaf developmental regulators often affects cotyledon development as well (Aida et al. 1997; Hanson et al. 2001; Qiu et al. 2002). They are also commonly used to study photomorphogenesis and skotomorphogenesis, light- and dark regulated development, respectively (von Arnim et al. 1996; Leivar et al. 2008; Wu et al. 2015). Some $C_4$ species, e.g. *Haloxylon ammodendron*, *Salsola soda* and other species in the *Salsola* genera, have cotyledons that perform $C_3$ photosynthesis (Lauterbach et al. 2017; Li et al. 2015; Pyankov et al. 2000). The reason for this is unknown, although it is thought to reflect the evolutionary development of $C_4$ photosynthesis (Pyankov et al. 2000). Others hypothesize that the moderate climate conditions, during which germination and the first days of seedlings development occur, favour $C_3$ photosynthesis (Lauterbach et al. 2017). In Bienertia, cotyledons contain the same tissues as true leaves and perform $SCC_4$ photosynthesis (Boyd et al. 2007). Microscopic observations of Bienertia seedlings have shown that the photosynthetic tissue in cotyledons develop the same morphology as the chlorenchyma cells of mature leaves (Akhani et al. 2005). Therefore, they are suitable for studying the development of $SCC_4$ morphology.

## 1.7 Skotomorphogenesis and de-etiolation

Plants grown in darkness undergo skotomorphogenesis, a process which channels the limited available nutrient resources toward rapid elongation of the hypocotyl in order to gain light exposure. The cotyledons remain small and underdeveloped. Accordingly, a dark-grown seedling is a heterotrophic entity, in contrast to a de-etiolated plant which performs photosynthesis and is thus photoautotrophic (von Arnim et al. 1996). A comparison of global gene expression in Arabidopsis and *Oryza sativa* (rice) showed a lower correlation in dark- than in light-grown seedlings, which led the authors to conclude that skotomorphogenesis was established more recently in plants than photomorphogenesis (Jiao et al. 2005). This simplified view has been

challenged by various researchers (Chen et al. 2010; Mathews 2006; Seluzicki et al. 2017).

Without light, the proplastid cannot develop into a chloroplast, and instead differentiates into an etioplast. The main function of the etioplast is the accumulation of the chlorophyll precursor, protochlorophyllide, so that upon illumination, chlorophyll is readily formed and photosynthesis can take place. Characteristic for an etioplast is the formation of the prolamellar bodies, a structure made of symmetrically arranged membranes that contain protochlorophyllide a. Upon light treatment, the prolamellar body is rapidly degraded, chlorophyll synthesis and thylakoid development starts and photosynthetic genes are induced, so that the seedling is photoautotrophic within hours (Wellburn et al. 1979; Tanaka et al. 1985). Etioplasts are mainly used to study the de-etiolation process, especially in regard to the assembly of photosynthetic complexes (Kleffmann et al. 2007; Kanervo et al. 2008). Etioplast-chloroplast differentiation leads to massive changes, but not of import machinery (Reiland et al. 2011). Etioplast formation under natural conditions was shown to occur in deeply sown seedlings (Whatley 1974), inner tissue layers of leaf buds (Solymosi et al. 2006) or cabbage heads (Solymosi et al. 2004).

## 1.8 Comparative transcriptomics and cross-species mapping

RNA-seq, a technique using next-generation sequencing (NGS), has recently been established as a method of choice for transcriptome analysis, surpassing microarrays in depth and coverage (Wang et al. 2009). It is especially popular for studies on non-model species, often in combination with *de novo* assembly of a transcriptome (Barrero et al. 2011; Lulin et al. 2012; Sun et al. 2012; Zenoni et al. 2010). However, the information gathered from these experiments is limited to the examined species. Comparative transcriptomics between species can be used to answer a wider range of questions. A comparative approach is not limited to RNA sequencing but also includes microarrays and meta-analysis of previously published studies. It has been used to identify genes essential for virulence by comparing pathogenic and non-pathogenic *Listeria* species (Wurtzel et al. 2012), study fistular leaf development in *Allium* (Zhu et al. 2017) and characterize the defense response of a *Fusarium* wilt disease-resistant tung oil tree (Chen et al. 2016). Cross-species comparison has proved to be especially useful in the case of $C_4$ photosynthesis. Comparative transcriptomics

of closely related $C_3$, $C_4$ and $C_3$-$C_4$ intermediate species have identified several new genes involved in $C_4$ (Furumoto et al. 2011; Bräutigam et al. 2011; Ding et al. 2015), regulatory mechanism controlling the development of Kranz anatomy (Yu et al. 2015), variations in $C_4$ metabolism (Covshoff et al. 2016) and given insight into the evolution of $C_4$ (Mallmann et al. 2014; Rao et al. 2016). The analysis of cross-species comparison can be conducted in two ways: a) a gene homolog-based analysis of separate single species data b) direct comparison of two or more species by using a single species microarray or transcriptome as a reference. The former method requires transcriptomes for all examined species as well as the identification of homologous genes. An advancement of this method was developed by Wang et al. 2014. In a study comparing the leaf gradient of maize and rice, thousands of gene homologs in both species, so-called anchor genes, were defined to normalize leaf developmental between the species. Using this method, considerable time and effort will be spent on identifying homologous genes. However, due to technical and time constraints, this approach is not feasible for all researchers. For the latter method to work, the compared species should be closely related and the effects of sequence mismatching have to be taken into consideration (Lu et al. 2009). With microarrays, this problem can be avoided by disregarding probes with a mismatch between species (Kirst et al. 2006) or using a multi-species array (Oshlack et al. 2007). With RNA-seq data, a mapping bias can be avoided by using specialized read mapping tools like BLAT (Kent 2002; Bräutigam et al. 2011).

## 1.9 Plant species used for comparison

**Suaeda heterophylla**

*Suaeda heterophylla* (Suaeda) is, like Bienertia, a halophytic plant of the subfamily Suaedoideae. Unlike Bienertia, Suaeda performs $C_3$ photosynthesis (Schütze et al. 2003). For this reason, it was chosen as an appropriate comparison species in this study. Apart from phylogenetic and anatomical features, little is known about this species.

*Beta vulgaris*

*Beta vulgaris* is an important crop plant. In 2014, 269.7 million tonnes of sugar beet were produced worldwide (http://www.fao.org/faostat). In the same year, its genome was sequenced and published. *B. vulgaris* belongs to the Amaranthaceae family like Bienertia and Suaeda and represents the first sequenced plant from this family (Dohm et al. 2014). Like Suaeda, *B. vulgaris* performs $C_3$ photosynthesis.

## 1.10 Aim of thesis

In this thesis, a comparative transcriptomics approach is used to investigate dark-grown cotyledon development in the $SCC_4$ species Bienertia and the closely related $C_3$ species Suaeda. The aim of this approach is to identify factors involved in $SCC_4$ development, in particular in regards to chloroplast positioning. As previous studies have shown that the cytoskeleton plays a role in $SCC_4$ development, a particular focus will be on finding cytoskeleton-interacting or -regulating factors. Furthermore, the identification of transcriptional regulators of $SCC_4$ development is anticipated. In addition, this thesis examines several cross-species mapping strategies to develop a workflow for comparative transcriptomics between the closely related species Bienertia and Suaeda.

# 2 Material and Methods

## 2.1 Materials

### 2.1.1 Chemicals and consumables

All chemicals and consumables were provided by either of the following companies, exceptions are noted in the text. All chemicals were at least of p.a. quality. AppliChem (Darmstadt, Germany), Bio-Rad (Hercules, USA), Biozym (Hessisch Oldendorf, Germany), Carl Roth (Karlsruhe, Germany), Duchefa (Haarlem, Niederlande), Eppendorf (Hamburg, Germany), Metabion (Martinsried, Germany), PEQLAB Biotechnologie GmbH (Erlangen, Germany), Promega (Madison, USA), Qiagen (Venlo, Niederlande), Roche (Basel, Schweiz), Sarstedt (Nürnbrecht, Germany), Sigma-Aldrich (St. Louis, USA), Thermo Fisher Scientific (Waltham, USA).

### 2.1.2 Kits

The kits used in this thesis are listed in Table 2.1.

Table 2.1: Kits used in this thesis

| Name | Company | Used for |
|---|---|---|
| GeneJET plasmid midi preparation kit | Thermo Fisher Scientific | Plasmid isolation |
| pENTR™/D-TOPO® | Thermo Fisher Scientific | Gateway cloning |
| Gateway® LR clonase® | Thermo Fisher Scientific | Gateway cloning |
| Platinum® SYBR® Green qPCR Super-Mix | Thermo Fisher Scientific | qPCR |
| RNase-free DNAse Set | Qiagen | RNA extraction |
| RNeasy MinElute Cleanup Kit | Qiagen | RNA extraction |

## 2.1.3 Buffers and media used

The buffers and media used in this thesis are listed in Table 2.2.

Table 2.2: Buffers and media used

| Name | Ingredients | Amount |
| --- | --- | --- |
| Infiltration medium | Saccharose | 292 mM |
| | Murashige & Skoog (MS) basal salt | 0.88 % (w/v) |
| | MES buffer salt | 20 mM |
| | Acetosyringone | 0.2 mM, add after autoclaving |
| Propagation medium, pH 5.7 | MS salt | 0.44 % (w/v) |
| | Saccharose | 60 mM |
| | MES buffer salt | 10 mM |
| | NaCl | 10 mM |
| | Gelrite | 0.8 % (w/v) |
| | 1000x MS vitamin stock | 1 X, add after autoclaving |
| Phosphate buffered saline (PBS), pH 7.4 | NaCl | 130 mM |
| | Na2HPO4 | 7 mM |
| | NaH2PO4 | 3 mM |
| LB medium (liquid), pH 7.2 | LB broth low salt | 0.1 % (w/v) |
| | tryptone | 1 % (w/v) |
| | yeast extract | 0.5 % (w/v) |
| | NaCl | 0.5 % (w/v) |
| LB medium (solid), pH 7.2 | LB agar low salt | 0.1 % (w/v) |
| | agar | 1 % (w/v) |
| | tryptone | 1 % (w/v) |
| | yeast extract | 0.5 % (w/v) |
| | NaCl | 0.5 % (w/v) |

## 2.1.4 Primers

Primers were synthesized by Eurofins Genomics (Ebersberg, Germany) or Metabion (Martinsried, Germany). They are listed in Table 2.3.

Table 2.3: Primers used in this study. Gene ID refers to the sequence number within the Bienertia (Bs) and Suaeda (Sh) transcriptome, which was used as a template for primer design.

| Gene | gene ID | Primer stock ID | sequence (5′ to 3′) |
|---|---|---|---|
| **qPCR** | | | |
| EF2 | Bs_UN55598 | 963_fwd | CCCTGTTGTGTCCTTCCGTGAGA |
| | | 964_rev | GCAAGCCCATCTTCCAGTGGAC |
| | Sh_292 | 967_fwd | CTGTTGTCCGTGTTGCTGTGCA |
| | | 968_rev | GATGGAGTTCACCAGCACCAGC |
| PEPC | Bs_UN000469 | 81_fwd | GCTTTAGGACATTGCAGCGGTATG |
| | | 82_rev | GCTCCTTACGAACAACAGAGCGGTA |
| POR A | Bs_1641 | 580_fwd | ACACAAACACATTGGCTGGA |
| | | 581_rev | GCAGCCAGGGTACAGAGAAG |
| NAD-ME | Bs_541 | 675_fwd | CTCTCGCTACGCTGTCATCC |
| | | 676_rev | CTAGGACTGTCGGATTCTCGCA |
| | Sh_661 | 879_fwd | AGAAACTGAGGGGCTTAGGGAG |
| | | 880_rev | GGCCAACACCTTGTCTACCA |
| APX5 | Bs_11466 | 891_fwd | TATGATAGTGACGGGTGCAACGA |
| | | 892_rev | GGACAGCATCCATATCCATTGGC |
| | Sh_1632 | 897_fwd | GCGGAATCACACAAGAAAATGGCT |
| | | 898_rev | GCCGCATATCCATACTGCCC |
| GAST1 | Bs_1551 | 909_fwd | GGCCAAATCTATGCCACAGGG |
| | | 910_rev | ACCGTGGGTGGTTTGGTTG |
| | Sh_4719 | 913_fwd | CAAGGTGTCAGTTAAGCCGTAGG |
| | | 914_rev | GCCGTGTGTGGTTTGGTTGTA |
| cysteine synthase | Bs_4374 | 883_fwd | GTCTTGATTGAGCCCACCAGTG |
| | | 884_rev | TACCTTTGGCTGGATCGGTAAGG |
| | Sh_4786 | 889_fwd | CTTCAAGGAAGGTCTGCTGGTTG |
| | | 890_rev | CTGGGAAAGACGGCAACAATGA |
| hydroxy-proline rich protein | Bs_37888 | 677_fwd | CCGTGTGCAACCTGTTTCAG |
| | | 678_rev | AATGGGCACTGAAGTCTCGG |
| | Sh_348 | 703_fwd | CAAGGAGGAACAGTGGGACC |
| | | 704_rev | GACTGGGCGCCTATCAAAGT |
| SOS3-interacting | Bs_72960 | 947_fwd | TGGTCGAGCTGAAGAAACAAAGTG |
| | | 948_rev | CGCCTTGCCAAGTCCAAAGAA |
| | Sh_3003 | 951_fwd | AGGTGTCCTGTCAATTGATGCAGA |
| | | 952_rev | CCCTTGCCATGCCCAAATGAT |
| NRAMP3 | Bs_27510 | 915_fwd | CGTGCTGCAGTCGATCCAAAT |
| | | 916_rev | AGTGTCTTGAGAACAGGGCCA |
| | Sh_9409 | 921_fwd | GGCTAAATGTGCTGCAGTCGA |
| | | 922_rev | GTGTCTTGAGAACAGGGCCG |
| anthranilate synthase | Bs_8080 | 925_fwd | GTCAACCCAAGCCCTTATATGGG |
| | | 926_rev | AGCAAGAGGTCTGTTAGTGATCTTCC |
| | Sh_6375 | 927_fwd | ACGTGGACCCTATAGTGGTGG |
| | | 928_rev | CTGCAGGGTCACTATCAGCAAC |
| **Gateway cloning** | | | |
| CHUP1 | Bs_7 (partial) | 169_fwd | CACCATGATAGTTAGGGTAGGCTTG |
| | | 178_rev | TTCTTCAACTTCACTAAATCG |

### 2.1.5 Plant material

Bienertia plants used for transient transformation were grown in phytochambers (Johnson Controls, Milwaukee, USA) under a 16 hours light, 8 hours dark cycle at temperatures of 30 °C during the day and 18 °C during the night. Plants were watered every other day with water containing approximately 0.1 % (w/v) table salt and 0.1 % (w/v) Wuxal® topN fertilizer. Plants used for experiments were typically 2 to 5 months old.

## 2.2 Methods

### 2.2.1 Molecular methods

**Plant material for RNA-seq experiment**

Bienertia and Suaeda seeds were germinated in petri dishes on wet filter paper in the dark at room temperature. To ensure absolute darkness, the petri dishes were wrapped in aluminium foil and stored in a box. Cotyledons were harvested separately at four different timepoints: post-germination (TP1), 2 and 5 days post-germination (TP2, TP3), and after 48 hours of illumination after TP3. Harvesting for TP 1-3 took place in the dark with a green safe light as the only source of light. Upon separation of the cotyledons, the plant material was frozen in liquid nitrogen and stored at $-80$ °C until RNA isolation. Four biological replicates were prepared.

**Microscopy**

For microscopy of Bienertia cotyledons, one to two seedlings from every TP were fixed overnight in 3 % formaldehyde and subsequently embedded in 5 % agarose. Cross-sections of 45 µm to 50 µm from the upper third of the cotyledon were produced with a VT1000 S vibratome (Leica, Wetzlar, Germany) and stored in 1 U PBS buffer until microscopy. Microscopy of cotyledon cross-sections was performed with

a standard microscope (Olympus CX31). Whole seedlings of Bienertia and Suaeda were photographed using a stereo microscope (Olympus SZX7).

## RNA isolation

Harvested cotyledons from all four replicates and timepoints of both species were pooled and ground in liquid nitrogen. Approximately 60 mg of ground plant material were dissolved in 1 ml Trizol and shaken vigorously for 15 min. 200 µL chloroform was added and the samples shaken for 10 min. The phases were separated by centrifugation (16 100 g at 4 °C for 15 min) and 400 µL of the aqueous phase were transfered to a new reaction tube. 99 µL of chloroform were added and the samples were centrifuged again (16 100 g at 4 °C for 15 min). 800 µL ice-cold 96 % (v/v) ethanol was added, the samples were incubated at −20 °C for 30 min and subsequently centrifuged (16 100 g at 4 °C for 15 min) to precipitate the RNA. The pellet was washed twice with ice-cold 70 % (v/v) ethanol and dissolved in 30 µL of ultrapure water (PURE-LAB Ultra, Elga LabWater, High Wycombe, UK). To degrade remains of genomic DNA, the RNA was treated with the RNase-Free DNase Set (Qiagen, Venlo, Netherlands) and concentrated with the RNeasy MinElute Cleanup Kit (Qiagen, Venlo, Netherlands) by eluting twice in 14 µL RNase-free water. Both kits were used following the protocols provided by the manufacturer. RNA concentration as well as 260/280 and 260/230 ratios were measured photometrically (Gen5 Take3™ Multivolume Plate with Synergy™ MX Microplate Reader, both Biotek, Highland Park, USA) and RNA integrity was checked by agarose gel electrophoresis. The RNA was stored at −20 °C until further use.

## Next-generation sequencing

Two biological replicates of TP1 to 3 from both species and TP4 from Bienertia were used for RNA-seq. 2 µg RNA per sample was sequenced with Illumina HiSeq2000 instruments following quality assessment, poly-A enrichment, cDNA synthesis with random hexamer priming and library preparation (GATC Biotech, Konstanz). At least 30 000 000 50 bp single reads per sample were provided in FASTQ format. The data for samples requiring multiple sequencing runs was concatenated for future proceedings.

## cDNA synthesis

500 ng RNA were used for cDNA synthesis. 50 pmol random primer was added to the samples, incubated at 70 °C for 5 min and immediately placed on ice. 1 mM deoxynucleotide triphosphates (dNTPs) (Thermo Fisher Scientific, Waltham, USA) and 100 U Moloney Murine Leukemia Virus Reverse Transcriptase (MMLV-RT) (Promega, Fitchburg, USA) were added and incubated at 37 °C for 60 min, followed by inactivation of the RT by incubating the samples at 70 °C for 10 min.

## qPCR

Quantitative PCR was carried out using a StepOnePlus cycler (Thermo Fisher Scientific, Waltham, USA) and Platinum® SYBR® Green qPCR SuperMix-UDG (Thermo Fisher Scientifc, Waltham, USA). cDNA dilution of the template was adjusted to fall within the standard curve. Bienertia and Suaeda gene homologs were identified with reciprocal BLAST and sequence alignment using the CLC Workbench 5.5 (CLC bio, Aarhus, Denmark). With the exception of the primers for *POR A* and *PEPC*, which have been previously designed for Bienertia and used for quality control of the RNA submitted for sequencing, primers were designed separately for each species. Primers were designed to have a melting temperature of 60 °C and produce an amplicon between 100 bp to 200 bp. After an initial denaturation step (2 min at 50 °C and 95 °C for 2 min), 40 cycles of 95 °C for 15 sec and 60 °C for 60 sec followed. At the end of the program, a dissociation curve was recorded (incremental temperature increase from 60 °C to 95 °C) to identify nonspecific PCR amplification and primer dimers. Quantification of transcripts was calculated with the StepOnePlus™ Software (Thermo Fisher Scientific, Waltham, USA) using the standard curve method. Elongation factor 2 (EF2) was chosen as a reference, as it was expressed at similar levels in samples of both species. All expression values are shown as relative to EF2.

## Amplification of BsCHUP1

The first 1044 bp of Bienertia *CHUP1* were amplified from cDNA by PCR using Phusion polymerase (Thermo Fisher Scientific) (Tables below). The primers used are listed in Table 2.3. The PCR product was examined for correct size by agarose gel

electrophoresis and subsequently purified using the QIAquick Gel Extraction Kit (Qiagen, Venlo, Netherlands).

| reaction setup | | amplification program | | |
|---|---|---|---|---|
| Phusion buffer | 1 X | 30 sec | 98 °C | |
| dNTPs | 200 µM | 10 sec | 95 °C | |
| Primers | 0.5 µM each | 10 sec | 58 °C | |
| Phusion | 1 U | 15 sec | 72 °C | 35 cycles |
| Template | variable | | | |
| Water | to 50 µL | 10 min | 72 °C | |

**Gateway cloning**

The Gateway® cloning system (Thermo Fisher Scientific) is an efficient way to insert DNA sequences into plasmid vectors. The procedure is carried in 2 steps. First, the amplicon is transferred into the entry vector (here, pENTR™/D-TOPO®), where it is flanked by a recombination sequence (attL). The destination vector (here, pEARLEY103) contains the matching recombination sequence (attR). pEARLEY103 is a vector containing a 35S-Promotor and a Green Fluorescent Protein (GFP) sequence outside of the recombination sequence so that GFP is fused to the C-terminal end of the amplicon (Earley et al. 2006). With the Gateway® LR reaction, the amplicon is transferred from the entry to the destination vector. The cloning procedure was performed according to the manufacturer's instructions. The vector plasmids containing the desired construct were amplified using transformed E. coli cells. Both plasmids were submitted for sequencing to ensure that the desired DNA sequence was inserted correctly (Seqlab, Göttingen). The sequencing results were compared to the Bienertia gene sequence using CLC Workbench 5.5 (CLC bio, Aarhus, Denmark).

*E. coli* **transformation**

Competent *Escherichia coli* (*E. coli*) DH5α cells were mixed with 100 ng to 200 ng plasmid and incubated on ice for 30 min. The cells were heat-shocked at 42 °C for 45 sec and immediately returned to the ice. The transformed cells were mixed with LB medium and incubated at 37 °C for 60 min on a shaker. The cells were plated on LB-agar containing the 50 µg ml$^{-1}$ kanamycin, and incubated at 37 °C overnight. Individual

colonies were picked and incubated in LB medium at 37 °C overnight. Plasmids were isolated using the GeneJet Plasmid Midiprep Kit (Thermo Fisher Scientific, Waltham, USA) following the manufacturer's instructions.

**Agrobacterium transformation**

The *Agrobacterium tumefaciens* (Agrobacterium) strain GV3101 (Koncz et al. 1986) was used for transient transformation of Bienertia. Competent Agrobacterium cells were mixed with 1 µg plastid DNA and incubated on ice for 5 min and subsequently transferred into a electroporation cuvette (Peqlab, Erlangen, Germany). A pulse of 2.18 kV/cm$^2$ was applied and the Agrobacterium/plasmid mixture was diluted in LB medium and incubated at 28 °C for 60 min. The transformed cells were plated on LB-agar plates containing 50 µg/ml kanamycin, 30 µg/ml gentamycin, and 50 µg/ml rifampicin, and incubated at 28 °C for 48 h.

**Transient expression in Bienertia**

Cuttings from Bienertia plants were transformed by Agrobacterium-mediated transformation. For this, the cuttings were inserted into a 50 mL falcon tube containing 30 ml infiltration medium with recombinant Agrobacterium cells grown to a OD of 0.4. Vaccum infiltration was carried out in 3 to 5 cycles of 10 sec at 850 mbar to 900 mbar, until the uptake of infiltration medium into the leaves was visible. The cuttings were incubated at room temperature and ambient light for 4-5 days in boxes containing propagation medium before extracting chlorenchyma cells for microscopy. Microscopy was performed with an Olympus CX31 microscope with a fitted fluorescence light source (U-LH50HG, Olympus).

## 2.2.2 Bioinformatic methods

**Reference transcriptomes**

The Bienertia transcriptome was assembled from a combination of 454 and Illumina reads from young and mature leaf tissues. The Suaeda transcriptome was assembled

from 50 bp Illumina reads from dark-grown cotyledons produced in this study. Both transcriptomes were assembled by Dr. Rick Sharpe, who kindly made them available for this study. The *B. vulgaris* transcriptome was downloaded from (Himmelbauer et al. 2017). The latest transcript assembly (BeetSet-2.genes.1408.mrna) was used as a reference.

### Identification of gene homologs

Gene homologs were identfied by performing a reciprocal BLASTX search (Altschul et al. 1990) with an e-value cutoff of $1 \times 10^{-6}$. If the sequences did not contain an ORF, a TBLASTX search was carried out instead. When the reciprocal BLASTX failed, an alignment was performed with the sequences identified in the reciprocal BLAST and percent identity was calculated to identify the closest match. If sequences aligned similarly well, all were retained and treated as gene homologs.

### Gene annotation

### Annotation of *B. vulgaris* transcriptome for MapMan use

To create a MapMan mapping file and gain additional access to functional annotation, the *B. vulgaris* transcriptome was annotated with the Mercator tool (Lohse et al. 2014). The annotation pipeline performs parallel sequence search against reference databases, including the reference organisms, Arabidopsis, Chlamydomonas, and rice, SwissProt annotation, and the protein domain databases InterPro, CDD and KOG. In addition, the software assigns a functional classification (MapMan BIN) to each gene using the MapMan ontology (Thimm et al. 2004). For some analyses, the full MapMan BINs assigned to a gene sequence (e.g. PS.lightreaction.photosystemII) were reduced to the primary BIN (e.g. PS). The output was imported into the MapMan software to be used as mapping file.

### Annotation of protein domains

There are multiple programs available online that provide a bioinformatic toolbox for protein domain identification. The ProSite Scan (Sigrist et al. 2013) checks the amino

acid sequence against a collection of motifs. The default option "Exclude motifs with a high probability of occurrence from the scan" was disabled. ProtScale (Gasteiger et al. 2005, http://web.expasy.org/protscale/) was used to test for hydrophobicity using the Kyte-Doolittle scale and default options. To detect coiled-coil domains, the tool COILS (http://www.ch.embnet.org/software/COILS_form.html, Lupas et al. 1991) was used with the default parameters. Protein localization prediction was carried out with the tool TargetP (Emanuelsson et al. 2007).

## Read mapping

### Read trimming

Before and after read trimming, read quality was assessed using the FASTQC tool (Babraham Bioinformatics 2017). Read quality was deemed acceptable if the Per Base Sequence Quality module reported a pass. To remove contamination with the Illumina Adapter, read trimming was performed with the trimming module of the RobiNA software (Lohse et al. 2012; Bolger et al. 2014) using the following modules with default values: Adaptor clipper, leading trimmer, trailing trimmer and sliding window trimmer.

### Mapping with CLC

Trimmed reads were imported into CLC Genomics Workbench 9.5 (CLC bio, Aarhus, Denmark) and mapped with the RNA-Seq Analysis tool. For mapping to the Bienertia transcriptome, default (strict) mapping parameters were used. Read mapping to the *de novo* assembled Suaeda transcriptome was performed with both default (strict) and adjusted (loose) parameters (Table 3.2). Reads mapping equally well to several genes were disregarded. The read counts per gene were normalized to reads per kilobase per million mappable reads (RPKM) (Mortazavi et al. 2008).

### Mapping with BLAT

The reads were mapped onto mRNA sequences of *B. vulgaris* genome (Dohm et al. 2014) by BLAT 3.5 (Kent 2002) using the high performance cluster system at the Leibniz Universität Hannover. Both reads and the reference transcriptome sequences were

Table 2.4: Parameters used for read mapping with CLC Genomics Workbench. Bienertia reads were mapped with strict parameters to the Bienertia reference transcriptome. Suaeda reads were mapped with strict and loose parameter to the Bienertia reference transcriptome.

| mapping | mapping parameters | | | | |
| | mismatch | insertion | deletion | length fraction | similarity |
| --- | --- | --- | --- | --- | --- |
| strict | 2 | 3 | 3 | 0.8 | 0.8 |
| loose | 1 | 2 | 2 | 0.5 | 0.8 |

translated in six frames to protein for the mapping (options -t=dnax -q=dnax; Listing 2.1).

Listing 2.1: Example script submitted to the computer cluster to map reads of the sample BsR1ZP1 to the reference transcriptome with BLAT in protein space

```
#!/bin/bash -login
#PBS -N BsR1ZP1_blat_bvul
#PBS -M l.hagenau@botanik.uni-hannover.de
#PBS -m ae
#PBS -j oe
#PBS -l nodes=1:ppn=1
#PBS -l walltime=40:00:00
#PBS -l mem=500MB

# show which computer the job ran on
echo "Job ran on:" $(hostname)

# module load ....
module load blat

# Change to work dir:
cd /home/nhealiha/
blat ./contig_assemblies/refbeet1-2_mrna.fasta /bigwork/nhealiha/
    LargeFiles/trimmedfasta/Bs_R1_ZP1_all.fasta
-t=dnax -q=dnax /bigwork/nhealiha/LargeFiles/psl/Bvulgaris/Bs_R1_ZP1.psl
```

The output was analyzed with the BLAT module pslReps to keep only the best hits for each read. The final output file was parsed with a custom python script, kindly provided by Jean-Marie Droz, to extract the total counts (TC) for each reference sequence.

**Data normalization and quality assessment**

Read normalization, principal component analysis (PCA) and differential gene expression were performed using the DESeq2 package (version 1.12.4, Love et al. 2014) with the R statistical package (versions 3.2.2 and 3.3.1, R Core Team 2017). The following script was used to calculate the normalized read counts (NC) from the TC data file and perform PCA. Unless otherwise stated, all plots and further calculations are based on the NC for the *B. vulgaris* mapping and RPKM counts for the CLC mappings.

Listing 2.2: script used for data normalization and PCA

```
### input file needs to have genes in rows, samples in columns ###
Design = data.frame( row.names = colnames(countTable),
timepoint = as.factor(c(rep("TP1",2),rep("TP2",2),rep("TP3",2),rep("TP4
    ",2),
                        rep("TP1",2),rep("TP2",2),rep("TP3",2))),
species= as.factor(c(rep("Bs",8), rep("sh", 6))),
replicate = as.factor(c("r1","r2", "r1","r2","r1","r2", "r1","r2",
                        "r1","r2","r1","r2","r1","r2")))

timepoint = Design$timepoint
species = Design$species
ddsTC <- DESeqDataSetFromMatrix(countData=countTable, colData=Design,
    design=~ timepoint)

#### define custom model matrix, necessary due to varying number of
    timepoints in both species ###
m1 <- model.matrix(~  timepoint + species:timepoint, colData(ddsTC))
### remove last (empty) column ###
m1 <- m1[,-8]
### reduced model matrix for likelihood ratio test ###
m0 <- model.matrix(~ timepoint, colData(ddsTC))
ddsTC <- DESeq(ddsTC,  test="LRT", full = m1, reduced = m0)

### access normalized counts ###
normCounts <- counts(ddsTC, normalized=TRUE)

### rlog-transformation for PCA analysis ###
rld <- rlog(ddsTC, blind=FALSE)
plotPCA(rld, intgroup=c("timepoint", "replicate", "species"))
```

**Coefficient of variation**

The coefficient of variation (CV) is defined as the ratio from the standard deviation to the mean. As the standard deviation increases with the mean, the unitless CV can be used to compare data with different means, in this case gene expression ranging from 0 to 5 622 046 NC. To define the expression cutoff value, the CV was calculated for each gene from NC by dividing the standard deviation of TP1 to 3 by the mean of TP1 to 3 separately for each species.

$$CV_{TP1-3} = \frac{SD_{TP1-3}}{mean_{TP1-3}}$$

For the filter criteria used to identify genes of interest (section 3.5.3), the CV was additionally calculated for TP3 of Bienertia.

**Differential gene expression**

Differential gene expression was calculated separately for each species using the DE-Seq2 package. For gene profiling and enrichment analysis, the test for differential expression was carried out with default parameters. This identified all differentially expressed genes that showed a log2 fold change different from zero. Changes should be of a certain magnitude to be considered biologically significant (Love et al. 2014). Thus, for identifying genes of interest, more stringent test parameters were chosen and the dataset was tested for differentially expressed genes with at least a twofold increase in expression (`results(ddsTC, lfcThreshold = 1, altHypothesis = "greaterAbs")`). Log2 fold changes and adjusted p-values were calculated between TP1 and 2, TP 1 and 3, and TP 2 and 3 for each species and betweeen TP3 and 4 for Bienertia. In both cases, a log2 fold change greater than absolute |1|, an adjusted p-value smaller than 0.1, and a mean expression of at least 20 NC in either species were used as cutoff value for differential expression.

**Enrichment analysis**

**Gene profiling**

Three developmental stages were sequenced for both species and differentially expressed genes were identified as described in section 2.2.2. To discern between which TPs genes were differentially regulated, genes were sorted into gene profiles using a custom R script. Eight distinct profiles were possible within this experimental setup (Table 3.4).

**Enrichment test with PageMan**

Enrichment analysis is used to find functional classes of genes over-represented in a larger dataset (Huang et al. 2009). Here, genes following specific expression patterns (gene profile) were tested for enrichment against all genes present in the *B. vulgaris* transcriptome. A PageMan native input file was generated, which assigned the genes contained in the respective gene profile a value of 1 and all other genes a value of 0, as described in the PageMan manual. Enrichment analysis was performed with the PageMan module of MapMan 3.5.1 using Fisher's exact test with the default cutoff setting of 1 and false discovery rate control via Benjamini-Hochberg correction (Usadel et al. 2006).

**Clustering**

Hierarchical and *k*-means clustering were performed with MeV 4.9.0 (Multiple Experiment Viewer, Saeed et al. 2003). Hierarchical clustering was performed for genes involved in cell cycle regulation and vascular development to assess developmental synchronicity between both species. Hierarchical clustering is often used to compare gene expression between samples, as it provides easy-to-read visual correlations between large data sets (Eisen et al. 1998). The NC were transformed into Z-scores, which reflects the gene expression shape regardless of mean expression values, i.e. a gene that increases in expression from 10 NC to 100 NC has the same shape (and Z-score) as a gene whose expression increases from 100 to 1000 NC. *Z*-scores were

calculated by gene within each species with the following formula:

$$z_{TPi} = \frac{NC_{TPi} - mean_{TP1-3}}{SD_{TP1-3}}$$

The Pearson correlation was used to calculate the distance matrix. Complete linkage was used as linkage clustering method, as it was shown to outperform average linkage (Gibbons et al. 2002).

*k*-means clustering was carried out for all differentially expressed genes tested for a log2 fold change of at least |1| (Section 2.2.2). To prepare the data for clustering, mean NC were normalized across genes using the Adjust Data menu within MEV. Clustering was carried out with 10 clusters, 50 maximum iterations and Pearson correlation as distance matrix.

### Comparison of sequencing data with Genevestigator

Genevestigator is a tool that integrates multiple microarray and RNA-seq experiments to allow for comparisons in gene expression across species, tissues and conditions (Hruz et al. 2008). Comparative studies were chosen with the Data Selection tool. For the comparison of light-induced photosynthetic genes, the dataset AT-00002, which contains data of 2 and 5 day old Arabidopsis seedlings grown in light and dark conditions, was selected. Arabidopsis gene homologs of *B. vulgaris* genes sorted into the primary MapMan BIN "PS" were identified by BLAST search and added to the Gene Selection in Genevestigator. The expression values from the chosen dataset were then exported and light-induction was calculated by forming the ratio between gene expression in light and dark-grown plants. If the ratio was higher than 2 in either the 2 or 5 day old seedlings, the gene was considered to be light-induced.

### Identification of genes of interest

### Candidate genes

The differential gene expression test for a log2 fold change greater than |1| was used as basis for identifying candidate genes. Genes were considered to be candidates if

they met the following criteria:

1. log2 fold change greater than 1, an adjusted p-value smaller than 0.1 between TP1 and 2, TP1 and 3, and TP2 and 3 in Bienertia
2. not identified as differentially expressed in Suaeda
3. a log2 fold change lower than 1 between TP3 and 4 in Bienertia
4. a CV lower than 40 % at TP3
5. a mean expression greater than 100 NC
6. a higher mean expression in Bienertia than Suaeda

The remaining genes were plotted and inspected individually. Genes with a similar expression profile in both species and comparable mean expression were excluded.

**Transcription factors**

Transcription factors potentially involved in $SCC_4$ development were identified by merging 3 data filtering approaches.

1. Transcription factors identified with the filtering criteria listed above
2. Transcription factors upregulated early in Bienertia (gene profiles 1+2) and not differentially expressed in Suaeda, with a mean expression at least twice as high in Bienertia (based on the dataset used for gene profiling (Section 2.2.2))
3. Transcription factors at least 5 times more abundant across TP1 to 3 in Bienertia compared to Suaeda

# 3 Results

## 3.1 Experimental design

Unpublished data on dark-grown cotyledon development of Bienertia has suggested that $SCC_4$ morphology develops in a light-independent manner, which preliminary microscopy confirmed. Thus, cotyledons were sampled at three developmental stages that reflect the progress of plastid arrangement from random to full $SCC_4$ morphology in order to find the genes responsible for the unusual organelle arrangement. Cotyledons of the closely-related $C_3$ species Suaeda were also sequenced to provide a "typical" developmental background from which the $SCC_4$ "factors" could more easily be identified. A comparison of both developmental transcriptomes should thus reveal the genes involved in $SCC_4$ morphology development.

## 3.2 Microscopy of Bienertia cotyledons

Three time points (TP1 to 3) of dark-grown cotyledons that represent different stages of morphology development (0, 2 and 5 days after germination) as well as an additional time point after exposure to light (TP4) were chosen for the experiment.

**Seedling development**

Just after germination, the cotyledons are curved and pale yellow green (Figure 3.1A and 3.1E). Most of time, they are still stuck inside the seed coat. Two days later, the cotyledons are longer and mostly straight. In about half of the samples, they appear reddish, most likely due to anthocyanine accumulation, while the others are yellow-green. Five days after germination, the apical hook is no longer apparent

and in some cases, the cotyledons have opened slightly. After a two-day exposure to light, the cotyledons are fully separated and most are green. However, in the case of previously red-coloured cotyledons, this process did not always fully occur within this time frame.
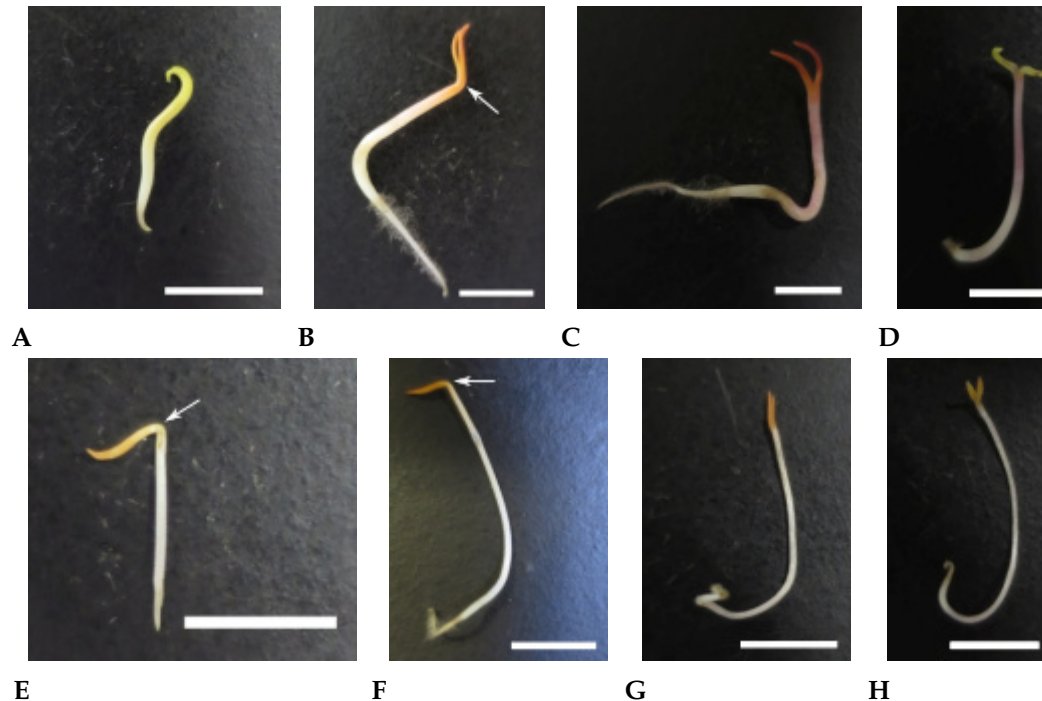


Figure 3.1: Dark-grown seedling development in Bienertia and Suaeda. Representative pictures of Bienertia (**A-C**) and Suaeda seedlings during development in the dark (**E-G**) and light (**D+H**). The arrow indicates the apical hook. **A+E**: TP1, dark-grown seedling post-germination. **B+F**: TP2, dark-grown seedling two days after germination **C+G**: TP3, dark-grown seedling five days after germination. **D+H**: TP4, seven day old seedling exposed to light for 48h. Scale bar = 5mm.

**Chlorenchyma cell development**

Microscopic analysis of various developmental stages of the expanding Bienertia cotyledon under light and dark conditions revealed the gradual establishment of the CC and positioning of the PCPs. Just after germination, light microscopy of a cotyledon cross-section shows two layers of chlorenchyma cells below the single epidermis layer (Figure 3.2A). Larger water-storage cells and developing vascular tissue is visible beneath the chlorenchyma. Plastids are visible in all tissue types, however, they are distributed randomly throughout the cell. The chlorenchyma cells contain a large nucleus (3.2B). Two days after germination, the CC is beginning to form in

the chlorenchyma cells. The chlorenchyma tissue is clearly distinguishable from the epidermis above and the water-storage cells below by the pale green color, caused by the now larger number of plastids. Plastids are accumulating around the nucleus, as well as lining the cell periphery (Figure 3.2D). However, they are more abundant and smaller than the plastids in mature cells and they do not show the typical donut-like shape of PCPs. The plastids in the surrounding tissues have a similar appearance (Figure 3.2C). The diameter of the hydrenchyma cells has expanded considerably, however, chlorenchyma cell size did not change significantly. Five days after germination, the CC is fully formed and clearly delineated, either positioned in the middle of the cell or at one side, while peripheral plastids are lining the cell walls (Figure 3.2F). Exposure to light for another two days leads to greening of the cotyledons, but no additional changes in cell morphology. There was no change in chloroplast positioning in the surrounding tissues (Figure 3.2G). In summary, $SCC_4$ morphology fully develops in dark-grown cotyledons of Bienertia within five days. It can thus be inferred that this process is regulated light-independently.

## 3.3 RNA sequencing and quality control

RNA extracted from cotyledons was prepared for Illumina sequencing for TP1 to 4 for Bienertia and TP1 to 3 for Suaeda (two replicates each). This produced a total of 624 415 619 and 247 978 802 50 bp single reads for Bienertia and Suaeda, respectively (Table 3.1). After quality control and processing, 617 906 121 trimmed reads for Bienertia and 247 000 354 trimmed reads for Suaeda were used for read mapping.

Table 3.1: Results of the Illumina sequencing runs. Raw reads are the total numbers of reads returned per sample, trimmed reads are after quality control and adapter trimming with RobiNA.

| | number of trimmed reads | | | |
| --- | --- | --- | --- | --- |
| | Bienertia | | Suaeda | |
| Time point | Replicate 1 | Replicate 2 | Replicate 1 | Replicate 2 |
| TP1 | 41 855 227 | 47 233 516 | 52 928 126 | 36 526 740 |
| TP2 | 74 442 990 | 60 834 201 | 51 939 452 | 33 116 999 |
| TP3 | 36 077 718 | 56 349 012 | 34 957 929 | 37 531 108 |
| TP4 | 48 372 042 | 44 810 722 | NA | NA |
| Σ trimmed reads | 409 975 428 | | 247 000 354 | |
| Σ raw reads | 415 019 592 | | 247 978 802 | |
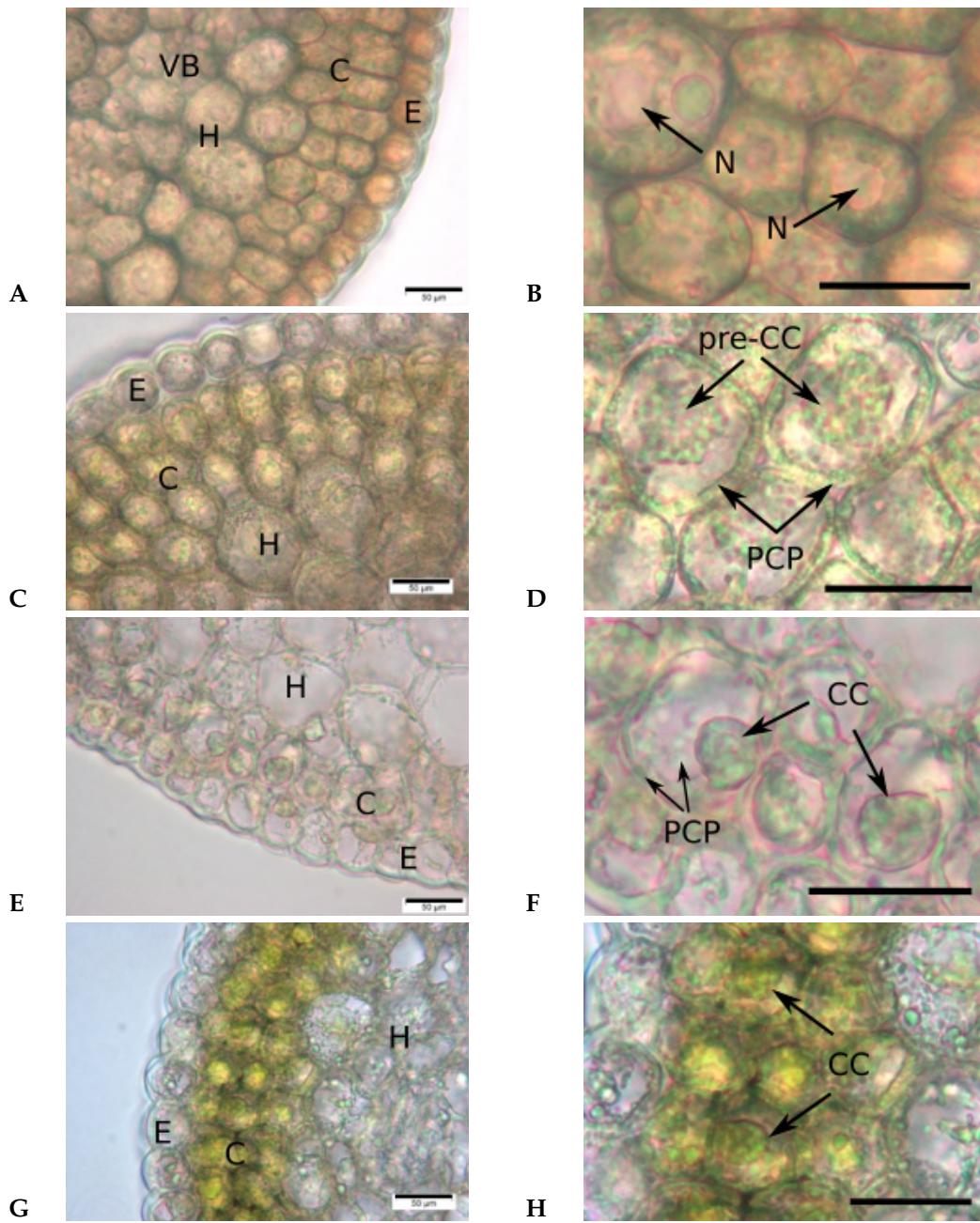
Figure 3.2: Light micrographs of Bienertia cotyledon and chlorenchyma development. Seedlings were germinated in the dark and sampled 0, 2, and 5 days post-germination (**A-E**) and after an additional 48 h exposure to light (**G+H**).**A+B**: TP1, post-germination. **C+D**: TP2, two days postgermination. **E+F**: TP3, five days after germination. **G+H**: TP4, seven days after germination with 48h exposure to light. C = chlorenchyma, E = epidermis, H = hydrenchyma, VB = vascular bundle, N = nucleus, pre-CC = pre-central compartment, PCP = peripheral plastid, CC = central compartment. Scale bar = 50μm

### 3.3.1 Read mapping

As the sequencing reads originated from two different species, it was important to choose the appropriate mapping procedure and reference transcriptome. Although both species are closely related, the differences in the gene sequences can introduce a mapping bias, if the read mapping is not carefully controlled. This is because most mapping algorithms are designed to align reads to the species the reads originate from. For faster computing times, these algorithms thus allow mismatches, but not insertions or deletions (indels). To identify mapping bias and find a cross-species mapping strategy, three different options were explored: Firstly, mapping both species in nucleotide space to the previously assembled and annotated Bienertia transcriptome. Secondly, mapping both species to their own transcriptome, which in the case of Suaeda had to be de novo-assembled from the reads produced in this experiment and lastly, mapping in protein space to the related species *B. vulgaris*, which has a well-annotated transcriptome. After read mapping, the total counts for each gene were normalized to relative abundances to ensure comparibility between the samples. In the case of nucleotide mapping, the normalized counts were RPKM, while the protein space mapping in combination with the DESeq2 package (Love et al. 2014) for detecting differentially expressed genes yielded NC (Section 2.2.2). The read mapping results for all described mapping scenarios can be found in Table 1 of the electronic supplementary material.

Table 3.2: Mapping percentages for all performed read mapping scenarios. All reads were mapped to the *B. vulgaris* transcriptome in protein space using BLAT and to the Bienertia transcriptome in nucleotide space with CLC Genomics Workbench. Suaeda reads were additionally mapped to the Suaeda *de novo* assembly.

| | mapping percentage | | | | | | |
| | Bienertia | | | | Suaeda | | |
| mapping | TP1 | TP2 | TP3 | TP4 | TP1 | TP2 | TP3 |
|---|---|---|---|---|---|---|---|
| BLAT to *B. vulgaris* | 63 %±9 | 64 %±1 | 67 %±8 | 60 %±4 | 57 %±3 | 55 %±2 | 54 %±1 |
| CLC to Bienertia (default parameters) | 70 %±1 | 64 %±3 | 68 %±1 | 62 %±2 | 48 %±1 | 44 %±1 | 45 %±1 |
| CLC to Bienertia (loose parameters for Suaeda mapping) | 70 %±1 | 64 %±3 | 68 %±1 | 62 %±2 | 71 %±1 | 69 %±0 | 69 %±1 |
| CLC to Suaeda | NA | NA | NA | NA | 79 %±0 | 78 %±1 | 78 %±0 |

**Mapping to the *de novo* assembly**

Both the *de novo* assembled transcriptome of Suaeda and the Bienertia transcriptome were used as mapping reference for each species separately. Reads were mapped using the mapping tool of CLC Genomics Workbench software with default parameters. The average mapping percentage was comparable between both species, with a slightly higher value for Suaeda (Table 3.2). This is to be expected for read mappings back to a *de novo* assembly, as the reads used for the *de novo* assembly are identical to the reads being mapped. While the read mapping was successful, downstream analysis proved to be very difficult, as gene homologs of both species had to be identified for a cross-species comparison. A BLAST search of the Bienertia against the Suaeda transcriptome did not find a homolog for 34.6 % of the Suaeda genes, which made it hard to compare the gene expression across the two species. Additionally, the Suaeda transcriptome was not annotated, which is a difficult and time-consuming operation despite the many bioinformatic tools available. As a result, this approach was abandoned and alternatives were considered.

**Mapping to the Bienertia transcriptome with CLC**

Initially, the clean reads from both species were mapped with the CLC Genomics workbench mapping tool to the Bienertia transcriptome using the default mapping parameters. This resulted in a much lower mapping percentage for Suaeda compared with Bienertia (43-50 % and 60-70 %, respectively; Table 3.2). To compensate for differences between the Bienertia and Suaeda transcriptome and to ensure that both Bienertia and Suaeda reads mapped with the same efficiency, the mapping parameters for Suaeda (mismatch, insertion, deletion, length fraction, and similarity) were relaxed step-by-step until the same mapping percentage was achieved. Relaxing the mapping parameters, however, increases the chance of mismatches to the transcriptome. The histogram of the mean of normalized reads for all samples per species demonstrates the impact of relaxing mapping parameters for both scenarios (Figure 3.3A). For the mapping with strict parameters, the Suaeda distribution appears skewed towards low expression values, i.e. the specificity is high, but the normalization procedure cannot compensate for the low efficiency. With loose mapping parameters, the Suaeda distribution is more symmetric, but has a narrower spread than the Bienertia distribution, resulting in at least twice as many counts for ex-

pression values between approximately 5 to 30 RPKM (Figure 3.3B). Compared to Bienertia, it is also shifted towards higher expression values. This pattern can likely be explained by mismatches in read mapping caused by the loose parameters, which lead to medium expression of multiple similar genes or isoforms. This reflects the non-specificity of the second approach.

**Mapping to the *B. vulgaris* transcriptome with BLAT**

As an alternative, both species were mapped against the recently sequenced *B. vulgaris* transcriptome (Dohm et al. 2014) in protein space using the Blast-like alignment tool BLAT (Kent 2002). The mapping efficiency was comparable for both species with 61 and 56 % of reads mapped to the reference transcriptome for Bienertia and Suaeda, respectively (Table 3.2). The histogram of mean gene expression shows that the distributions of both species are almost identical (Figure 3.3C). Thus, it can be concluded that there are only negligible differences in mapping for both species. A mapping bias likely still exists, however it appears to be directed equally against both species. A cross-species comparison should be possible and more reliable with this approach.



**A**  **B**  **C**
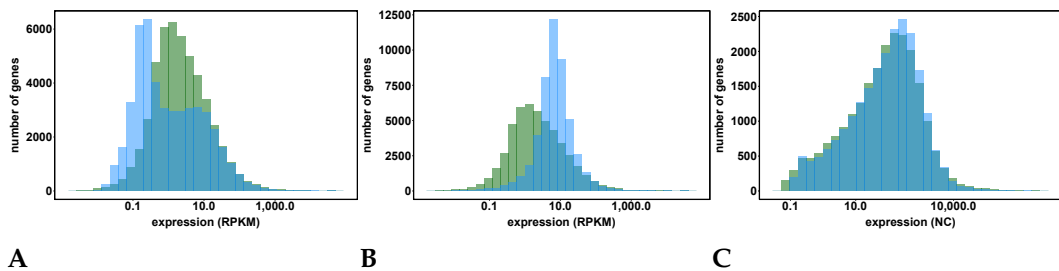
Figure 3.3: Histogram of read mapping data. The distribution of normalized reads by species was plotted to assess differences in distribution. Green represents Bienertia. Light blue represent Suaeda, Dark blue represents overlapping areas. **A**: mapping to the Bienertia transcriptome with strict parameters. **B**: mapping to Bienertia with loose parameters. **C**: mapping to *B. vulgaris* transcriptome with BLAT

### 3.3.2 Quality control

Before starting downstream analysis, it is important to ascertain the robustness and reliability of the data. As differences between genes with low read counts can often be

attributed to noise (Anders et al. 2010), a cutoff expression value was set and potential
sample outliers were identified.

**Principal component analysis**

Principal component analysis (PCA) was used to get an overview of the data and
identify potential sample outliers (Conesa et al. 2016). As such, it serves as a first
step in quality assessment before further analysis takes place. PCA reduces data
into so-called principal components, the first of which represents the largest possible
variance in the dataset. In general, replicates of the same sample cluster together.
This is observable in the PCA plot of the dataset (Figure 3.4), although some samples,
e.g. Bienertia TP4 and Suaeda TP2, cluster more loosely.
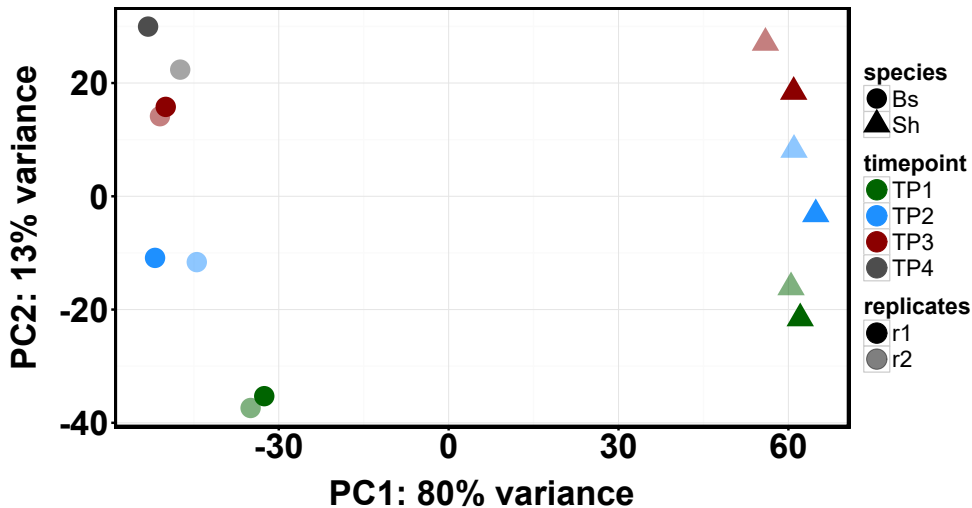


Figure 3.4: PCA plot of all samples. Dots represent Bienertia, triangles represent Suaeda.
Colors represent time points.

Here, the first PCA component separates the species, accounting for 80 % of the var-
iance, while the second component separates the time points, accounting for 13 % of
the variance. This suggests that differences in the data can mostly be attributed to dif-
ferences between the species and only secondly to the developmental stages. None of
the samples behaved like an outlier, thus, the data was deemed acceptable for further
analysis.

**Determination of the cutoff value**

Differential expression analysis of the data was performed with DESeq2 (Love et al. 2014), which determines the expression cutoff value on a model-based dispersion estimate. However, for some of the downstream analyses, non-differentially expressed genes were also of interest. Genes with low expression values have a high CV and are thus unreliably quantified. A cutoff value was needed to exclude these genes from further analysis. To define a cutoff expression value, the CV, defined as the ratio of standard deviation and mean of NC, was calculated for all samples of each species and plotted against the mean expression value (Figure 3.5).



**A**          **B**
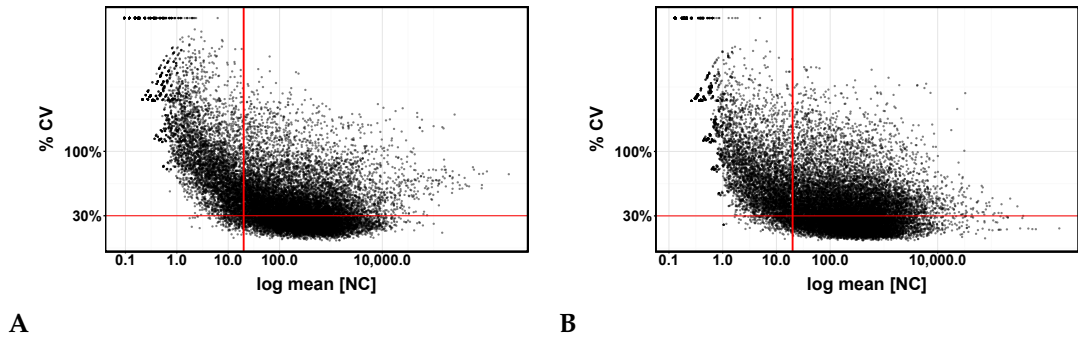
Figure 3.5: Determination of a low expression cutoff value for further analysis. The x-axis shows the mean of normalized counts (NC) as calculated by DESeq2 on a log-scale. The y-axis shows the CV in percent (% CV), defined as the ratio of standard deviation and the mean. The vertical red line intersects at the chosen cutoff value of 20 NC. The horizontal line shows the approximate average %CV for the values above the cutoff. **A**: Bienertia **B**: Suaeda.

The CV drops sharply as expression increases to approximately 20 NC, after which the decrease in CV is slower. At this point, the background noise of random reads no longer masks real gene expression. The lowest CV value of about 30 % is reached at approximately 100 NC for both species. For downstream analysis, a general cutoff of at least 20 NC was thus determined. For the highest expressed genes in Bienertia, the % CV increases again. This is uncommon behaviour, as the CV usually decreases with the expression. One possibility is that this is caused by read duplicates in one of the samples. Duplication levels were estimated by the FastQC tool (Babraham Bioinformatics 2017) and show a higher average duplication level for one of the Bienertia replicates. However, that is also the case for one of the Suaeda replicates, although duplication levels are generally lower for the $C_3$ species. Read duplicates can be either naturally occuring highly transcribed genes and thus of biological origin or PCR artifacts that arise during library preparation, especially when working with limited

starting material. It is not possible to distinguish naturally occuring read duplicates and PCR artifacts bioinformatically. It is also not recommended to remove duplicates from the data as it does not improve accuracy or false discovery rate (Parekh et al. 2016). However, none of the genes was flagged as outlier during DESeq2 analysis.

### 3.3.3 Mapman functional annotation

As the *B. vulgaris* transcriptome was not included in the mapping files provided by the Mapman software, the Mercator tool (Lohse et al. 2014) was used to annotate the transcriptome with Mapman BINs. Mercator searches multiple databases like InterPro, Arabidopsis TAIR 10, SwissProt, Uniref90, cdd, KOG and others to classify proteins or gene sequences into functional categories (BINs). Around 60 % of the transcriptome could be assigned to a BIN, while 40 % remained unassigned. The distribution of genes to a primary Mapman BIN is shown in Figure 3.6). The primary BINs protein, RNA, misc, signalling, stress and transport have over 1000 genes assigned to them. The BINs micro RNA, gluconeogenesis/glyoxylate cycle and S-assimilation have less than 10 genes assigned.
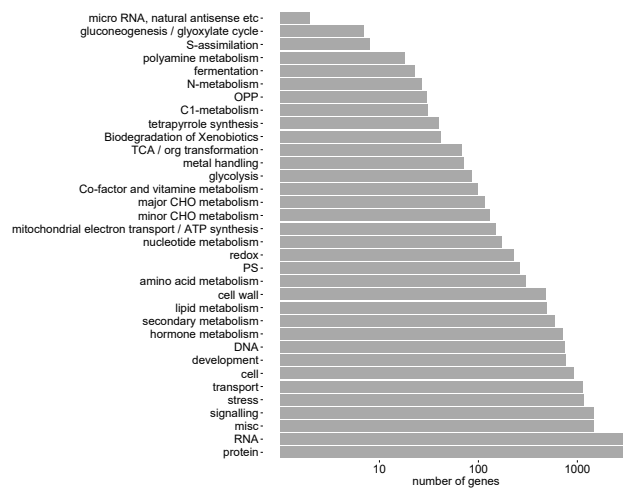


Figure 3.6: Distribution of primary MapMan BINs in *B. vulgaris* transcriptome. Functional annotation was carried out using the Mercator tool. Unassigned genes are not shown. The x-axis shows the number of genes sorted into the respective BIN.

### 3.3.4 Control for dark treatment

To enable cotyledon harvesting, the dark-grown seedlings had to be exposed to a green "safe light" for a short amount of time. There have been reports that even these wavelengths could trigger changes in gene expression (Zhang et al. 2012). The morphology of the seedlings, particularly hypocotyl elongation and closed cotyledons (Figure 3.1), as well as their pale colour suggest that they have not started de-etiolation. To confirm that neither of the dark-treated samples has started photomorphogenesis, the expression pattern of several genes known to be either repressed or induced by light were examined. The genes were selected based on literature and the *B. vulgaris* homologs were identified by BLAST search. These include the NADPH:protochlorophyllide oxidoreductase A (POR A), PEPC, elongated hypocotyl 5 (HY5), phytochrome A (Phy A) and constitutive photomorphogenesis 1 (COP1) (Figure 3.7). The expression patterns in light- and dark-grown Arabidopsis seedlings were extracted from literature as well as the eFP browser Light Series (Winter et al. 2007) and compared to gene expression in Bienertia and Suaeda. Expression of *POR A* and *PEPC* were also measured by qPCR, as TP4 was not sequenced in Suaeda.
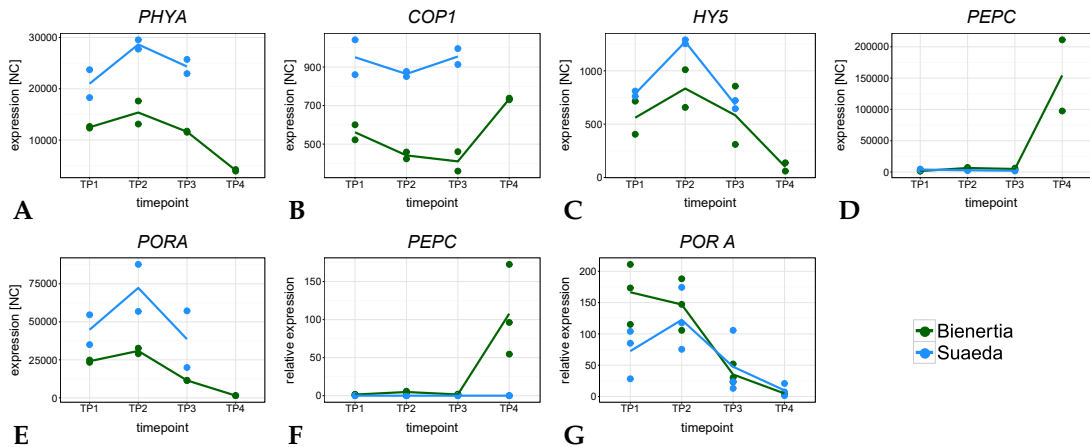


Figure 3.7: Expression analysis of light-regulated genes during cotyledon development in Bienertia and Suaeda. **A-E**: Expression according to RNA sequencing. The y-axis shows NC as calculated by the DESeq2 package. **F+G**: qPCR analysis of *PEPC* and *POR A*. The y-axis shows relative expression. The data was was normalized to the expression of EF2.

Phy A is one of the red-light photoreceptors that starts the signaling cascade for photomorphogenesis and it has been shown to be downregulated after exposure to light (Colbert 1991). Additionally, both mRNA and protein are unstable in light (Reed et al. 1994; Maloof et al. 2001; Colbert 1991). The eFP brower data also indicates downreg-

ulation after a 4 hour exposure to light. It is expressed highly in both Suaeda and Bienertia during TP1 to 3 and clearly downregulated at TP4 in Bienertia. COP1 is an E3 ubiquitin ligase that degrades positive regulators of photomorphogenesis in dark-grown plants and is involved in the degradation of both HY5 and Phy A. Its expression is not reported to change after exposure to light, instead, its activity is regulated by relocation from the nucleus to the cytoplasm (Yi et al. 2005). In Bienertia and Suaeda it is constitutively expressed at TP1 to 3, and slightly increases in expression after illumination. This corresponds to the eFP browser data. HY5 is a bzip transcription factor that is negatively regulated by COP1 in darkness. Its transcript abundance is two- to threefold higher in light-exposed than dark-grown seedlings according to literature and the eFP data (Osterlund et al. 2000). However, expression of *HY5* at TP4 in Bienertia is much lower than at TP1 to 3, showing the opposite pattern. Microarray data from Arabidopsis leaves grown in a 12 hour light cycle show downregulation of *HY5* after 4 to 8 hours of light. PEPC is one of $C_4$ core genes and previous experiments have shown that it is expressed highest in mature photosynthetic tissue in Bienertia, while almost non-existent in younger leaves (Lara et al. 2008). It is light-induced and its expression in darkened leaves is decreased (Lara et al. 2008). In $C_3$ plants, a form of PEPC is involved in glycolysis, but not photosynthesis. Both the RNA-seq and the qPCR data show the expected expression pattern for both species, i.e. low expression in the dark-grown cotyledon samples of Bienertia and Suaeda (TP1 to 3) and high expression in the light-exposed sample of Bienertia (TP4). POR A is an enzyme that catalyses the second-to-last step in chlorophyll synthesis and is known to be downregulated strongly during light exposure (Runge et al. 1996). Both species show high expression of *POR A* in the early dark-grown developmental stages and a sharp decrease is observed in the Bienertia cotyledons exposed to light. The qPCR data confirms the pattern observed in the RNA-seq data for both species.

Furthermore, light-induced photosynthetic genes (43 genes in total) in Bienertia were identified and compared to Arabidopsis data (AT-00002) with the Genevestigator tool (Hruz et al. 2008). The microarray dataset consisted of two and five day old Arabidopsis seedlings grown in dark and light conditions. 80 % of the light-induced genes in Bienertia overlapped with the Arabidopsis dataset (Figure 3.8). However, the total number of light-induced photosynthetic genes in Arabidopsis was much higher with 102 genes. Of the 68 photosynthetic genes uniquely light-induced in Arabidopsis, 39 are expressed at high levels (>1000 NC mean) in Bienertia, often with increasing expression levels over time (Table A.2). It is possible that 48 h of light exposure af-

ter five days of growth in the dark are not enough to reliably induce expression in Bienertia.
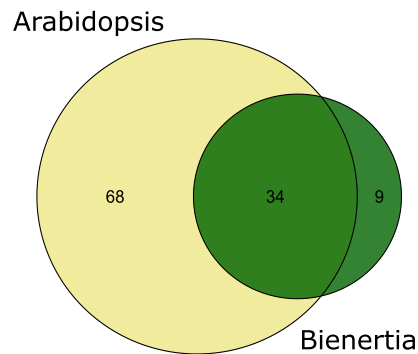


Figure 3.8: Comparison of light-induced photosynthetic genes in Arabidopsis and Bienertia. The size of the circles represents the total number of genes. The numbers represent the overlapping and uniquely occurring genes.

In summary, several known light-repressed or light-induced genes were examined and, with the exception of HY5, found to behave as expected in the dark- and light-treated samples, respectively. A larger scale observation of light-induced genes involved in photosynthesis as well as clear hypocotyl elongation in the seedlings similarly suggest that photomorphogenesis did not set in until exposure to light. It was hence deemed acceptable to analyse the data further under the premise of development in the dark.

## 3.4 Comparison of cotyledon development in Bienertia and Suaeda

### 3.4.1 Synchronicity of development

To ascertain that basic developmental processes occur at the same time and speed in both species, microscopic and cellular markers of cotyledon development were assessed. Microscopy of the seedlings in Bienertia and Suaeda show an increase in cotyledon size as well as hypocotyl and root growth from TP1 to 2 (Figure 3.1), but not from TP2 to 3. In both species, the apical hook is visible in early developmental stages, but disappears between TP2 to 3. While this is a first indication of similar

development, it is not sufficient. Next, the growth process was examined at a cellular level. In developing and growing tissue, cells first undergo proliferation by dividing mitotically. This is followed by cell expansion, a process that involves enlargement of the central vacuole. The latter step often occurs at the same time as endoreduplication, the process of replicating chromosomes without cell division. Mature cells are fully differentiated and have stopped growing. These steps are regulated by a large array of core cell cycle genes. Younger tissues contain more proliferating and expanding cells than mature cells which should be reflected by high expression of the cell cycle genes and regulators of cell expansion. Cotyledons of Bienertia and Suaeda contain several tissues, such as epidermis, hydrenchyma, photosynthetic and vascular tissue (Figure 3.2). Suaeda contains three to four layers of M cells as photosynthetic tissue and small hydrenchyma cells surrounding the vascular bundles. Bienertia has two layers of chlorenchyma cells and large hydrenchyma cells around the vascular network. With the differences in anatomy for the photosynthetic and water storage tissue, vascular development seems to be suitable to explore developmental synchronicity in the two species. Vascular development is a well-studied process of tissue differentiation in plants and several major regulating genes have been described. To test for synchronicity in development of the species, the expression of core cell cycle genes and genes responsible for vascular development were analysed.

**Cell cycle genes**

Arabidopsis gene IDs for the core cell cycle genes were obtained from Külahoglu et al. (2014), Beemster et al. (2005), and Vandepoele et al. (2002) and *B. vulgaris* homologs were identified using a combination of BLASTX and the Mercator mapping tool (Section 2.2.2, 2.2.2). 57 core cell cycle genes were thus identified in *B. vulgaris* and their expression pattern was explored in Bienertia and Suaeda using hierarchical clustering. Isoforms were identified in the *B. vulgaris* transcriptome for 17 of the genes. These were included in the dataset, for a total of 84 genes. Before clustering, z-scores were calculated from the mean expression values of each gene to center the data and make it suitable for comparison (Section 2.2.2).

The result is shown in Figure 3.9 and clearly demonstrates that the majority of the genes follow the same pattern in Bienertia and Suaeda. The genes are sorted into seven main clusters (A-G). Clusters A, B, D and E contain 60 genes (71 %) and represent

Figure 3.9: Hierarchical clustering of the core cell cycle gene expression over TP1 to 3 in Bienertia and Suaeda. The genes are normalized, with blue representing low expression and yellow high expression. The clusters marked green represent synchronous expression over time in both species. The clusters marked red show desynchronized expression between the species. The black-orange bar on the right side of the cluster shows functional annotation of the genes. Genes expressed during cell proliferation and positive regulators of the cell cycle are marked black, genes involved in cell differentiation and cell growth are marked orange.

44

similar expression patterns between the species. Clusters C, F and G contain 24 genes (29 %) and show different expression patterns between Suaeda and Bienertia. Genes involved in the cell cycle, like cyclins and cyclin-dependent kinases (CDK), as well as positive regulators of the cell cycle are represented by the black bar on the right side of the cluster. They make up the majority of the genes. Genes controlling cell differentiation and growth are marked orange. Clusters A and D contain 43 genes that peak in expression at TP1 and 2, respectively, in both species and are downregulated at TP3. With the exception of one gene, all genes are transcribed in proliferating tissue. Clusters B and E consist of eight and nine genes, respectively with expression peaks at TP2 and 3. Of these genes, seven are involved in cell differentiation and expansion. The majority of the genes involved in cell proliferation is thus expressed at TP1 and 2, but downregulated at TP3. Negative regulators of the cell cycle and genes involved in cell differentiation and cell growth are upregulated late. The expression patterns of 24 genes are not correlated between both species (clusters C, F and G). Genes in the cluster F show only a slight delay in gene expression in Suaeda compared to Bienertia, while 14 genes are highest expressed at TP1 in Bienertia, but are upregulated late in Suaeda (cluster G). Several cyclins and cyclin-dependent kinases (CDK) are in this cluster. The HOBBIT gene has been shown to regulate cell differentiation in *A. thaliana* (Blilou et al. 2002). CDK inhibitor 6 (ICK6) is one of the ICKs that are typically expressed to exit the cell cycle. DP-E2F-LIKE1 (DEL1) preserves the mitotic state of the cell and prevents endoreduplication (Vlieghe et al. 2005). The non-correlated genes could be responsible for differences between the two species by controlling the development of the photosynthetic and hydrenchyma tissue. Over 70 % of the core cell cycle genes share a similar expression pattern across both species. Overall, a progression from proliferation to maturity over the course of the experiment is observed.

**Vascular development**

17 genes responsible for vascular development were obtained from literature. Determination of *B. vulgaris* homologs and normalization of gene expression for clustering was performed as previously described. For three of the genes, several isoforms were identified in the *B. vulgaris* transcriptome. These were included in the dataset. Clusters B, C and D contain 13 genes (65 %) that are similarly regulated in both species. Cluster C mostly contains early regulators of vascular development (AS1, ATHB8,

KAN2). Genes involved in vein formation and vascular patterning are expressed at TP2 and 3 (cluster B+D). Seven genes (35 %) show different expression patterns in Bienertia and Suaeda. Interestingly, both isoforms of the auxin efflux protein PIN1 are upregulated late in Bienertia compared to Suaeda. PIN1 has been shown to be differentially regulated in $C_4$ plants compared to $C_3$ species due to the high vein density needed for Kranz anatomy (Slewinski et al. 2012). However, Bienertia as a $SCC_4$ plant does not have a higher vein density as its $C_3$ relatives (Freitag et al. 2002). The bHLH transcription factor LONESOME HIGHWAY, which is involved in xylem differentiation in A. thaliana (Ohashi-Ito et al. 2013), has two isoforms in *B. vulgaris*. Curiously, one of the isoforms is highly expressed in Bienertia and lowly expressed in Suada, while the other isoform shows the opposite pattern. The majority of genes involved in vascular development shows the same expression pattern in both species. Early regulators of vascular development are expressed highest at TP1 and downregulated at TP2 and 3, while most genes controlling vein patterning are upregulated later.
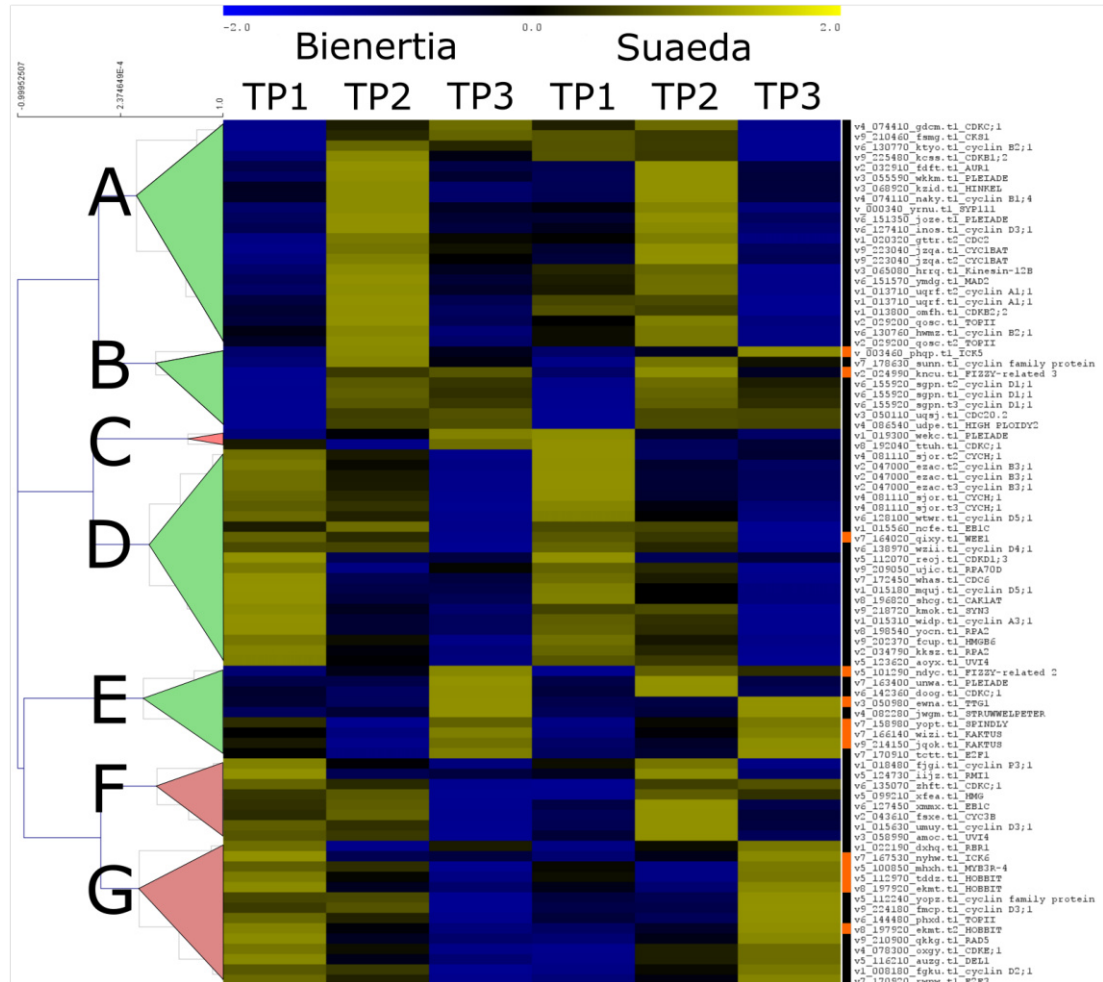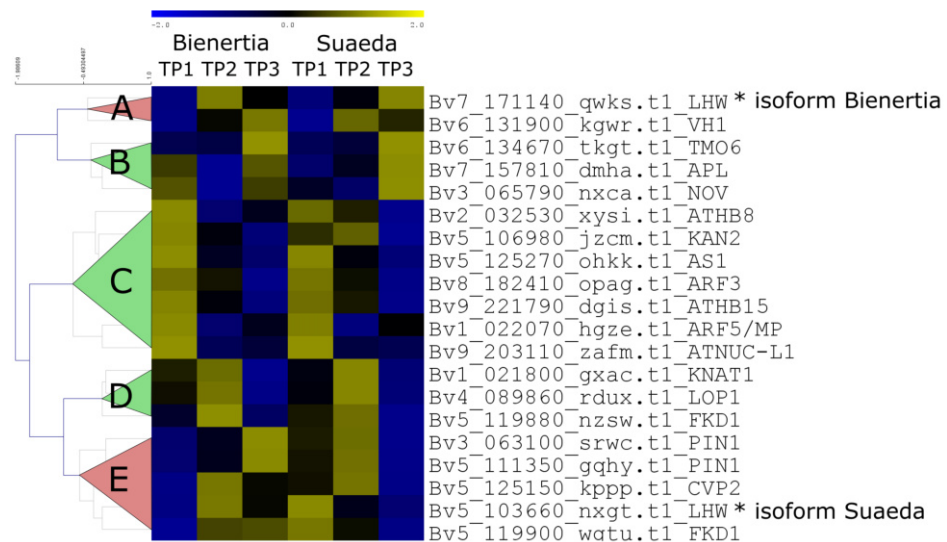


Figure 3.10: Hierarchical clustering of vascular development gene expression over TP1 to 3 in Bienertia and Suaeda. The genes are normalized, with blue representing low expression and yellow representing high expression. The clusters marked green show synchronous expression over time in both species. The clusters marked red show desynchronized expression between the species.

In summary, the largely simultaneous expression patterns of cell cycle genes and vascular development regulators point to concurrent development in both species. The macroscopic development of the seedlings also suggests that both species are undergoing similar overall changes. Although individual genes are not correlated

between Bienertia and Suaeda, large scale differences in gene expression can not be explained by a general lag of development in one of the species.

## 3.4.2 Temporal patterns of gene expression

An initial assumption was that the genes expressed during development as well as the changes in expression are mostly overlapping between the two species, as they undergo the same process of skotomorphogenesis. However, the PCA analysis suggests that most of the variation in the data stems from differences between the two species (Section 3.3.2). The variation in the data is further examined by looking at the overlap of genes at several expression levels between the species as well as gene expression patterns of differentially regulated genes. For the rest of the study, low expression refers to NC values between 20 and 100 NC, medium expression to 100-1000 NC and high expression to values over 1000 NC. The species-wise overlap between gene expression and expression strength is summarized in Table 3.3. As expected and previously shown (Figure 3.3), there are roughly the same number of genes at all expression levels in Bienertia and Suaeda. In Bienertia, 48 % of the lowly expressed genes are also lowly expressed in Suaeda (Table 3.3). For genes with a medium and high expression, the overlap increases to 77 and 74 %, respectively (71 and 72 % for the Suaeda/Bienertia ratio). The majority of non-overlapping genes are found in an adjoining category, e.g. 24 % of the genes that are highly expressed in Bienertia have a medium expression in Suaeda, while only 1 % are lowly expressed (Table 3.3). That means, the majority of genes have the same level of expression in both species with only a small percentage of genes differing drastically in terms of overall expression. However, these are broad categories and comparison of gene expression levels do not reflect the changes in gene expression over time.

**Transcriptional investment**

All genes above 20 NC were used to show the transcriptional investment at every developmental stage for both species (Figure 3.11). Transcriptional investment refers to the cumulative NC per primary Mapman BIN (Section 2.2.2). For TP1 and 2, the transcriptional profiles appear similar in both species. At TP1, 13 % to 14 % of gene expression is found in the BIN "protein", while photosynthetic genes make up 4 % of accumulated NC. At TP2, photosynthetic gene expression increases in both species

Table 3.3: Overlap of gene expression levels between Bienertia and Suaeda in percent. Low expression was defined as 20-100 NC as calculated by DESeq2, medium expression as 100-1000 NC and high expression as >1000 NC. Grey cells: percentage of genes with the same expression strength in both species.

|  |  | Bienertia | | |  |  | Suaeda | | |
|---|---|---|---|---|---|---|---|---|---|
|  |  | low | medium | high |  |  | low | medium | high |
|  | low | 48 % | 11 % | 1 % |  | low | 50 % | 20 % | 1 % |
| Suaeda | medium | 33 % | 77 % | 24 % | Bienertia | medium | 21 % | 71 % | 27 % |
|  | high | 0 % | 9 % | 74 % |  | high | 1 % | 8 % | 72 % |

to 9 % in Bienertia and 6 % in Suaeda. At TP3, the transcriptional profiles diverge to a certain degree between the species. There is an increase in unknown genes in both species. In Bienertia, it increases by 13 % compared to TP2, while in Suaeda, the increase is smaller (3 %). Gene expression in the "protein" category is reduced to 8 % in Bienertia and to 12 % in Suaeda. Photosynthetic gene expression is also reduced in both species compared to TP2, but higher in Bienertia than in Suaeda (7 % and 4 %, respectively). TP4 in Bienertia has a similar profile to Bienertia TP3, however, photosynthetic gene expression has increased to 10 %.
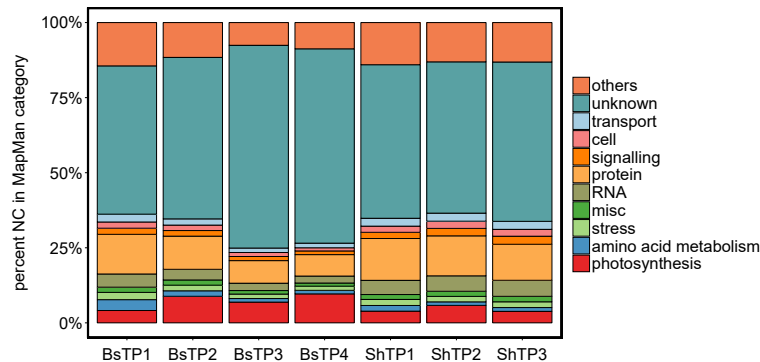


Figure 3.11: Transcriptional investment during cotyledon development in Bienertia and Suaeda over all time points. All genes above 20 NC were sorted into primary MapMan categories. The y-axis shows the added NC in percent for each category.

**Developmental patterns of gene expression**

To analyse the developmental patterns of gene expression in both species and whether these patterns overlap, the dataset was filtered for differentially expressed genes. The DESeq2 package in R (Love et al. 2014) was used to identify differentially expressed

genes between the developmental stages in both species. This analysis was performed separately for both species. 3269 genes in Bienertia and 6232 genes in Suaeda were identified as differentially expressed. This equals 21 % and 38 % of all expressed genes for Bienertia and Suaeda, respectively. This large difference in number can partly be explained by the slightly lower variance in the Suaeda samples which allows DESeq2 to identify small fold changes as differential expression. This is reflected in the number of differentially expressed genes with a log2 foldchange between -1 and 1, which is much higher for Sueda than Bienertia. A log2 foldchange cutoff value of 1 (twofold increase or decrease) was applied to the dataset for the following analyses. This corresponds to 2631 genes in Bienertia and 3821 genes in Suaeda. Each gene was then assigned to a profile to identify overall temporal patterns of gene expression (Table 3.4). The list of all differentially expressed genes can be found in Table 2 of the electronic supplemental material.

In both species, most genes have either an upward or downward trend (profiles 1, 2, 4 and 5, 7, 8, respectively), while there are few genes that show an up-down or down-up pattern throughout cotyledon development (profiles 3 and 6), corresponding to 4 and 3 % of the differentially expressed genes for Bienertia and Suaeda, respectively. 10 and 16 % of the differentially regulated genes are found in profiles 1 and 8, meaning they are consistently up- or downregulated during cotyledon development. The majority of differentially expressed genes are found in profiles 2, 4, 5 and 7. The ratio between genes with an upward versus downwards trend is 1.75 in Bienertia and 0.9 in Suaeda. This means that there are almost twice as many genes upregulated in Bienertia than downregulated, while the number of up- and downregulated genes is roughly the same in Suaeda. Of the upregulated genes in Bienertia, the majority are upregulated early, between TP1 and 2 (profile 2). For Suaeda, more genes are upregulated between TP2 and 3 (profile 4) than TP1 and 2 (profile 2). To test whether the expression of individual genes follows the same trend in both species, genes with an upward or downward trend were compared between both species. Less than half of the genes (42 %) with an upward trend in Bienertia also show also an upward trend in Suaeda, while the number is slightly higher for genes with a downward trend (54 %). The non-overlapping genes thus show either the reverse trend (5-7 %) or are not developmentally regulated at all. This means that most developmentally regulated genes follow a different expression pattern in the two species.

Table 3.4: Gene expression profiles for all differentially expressed genes. All differentially expressed (DE) genes from both species were sorted into developmental expression profiles. The profile sketches show all eight possible gene expression patterns for TP1 to 3.
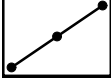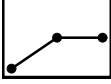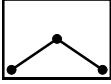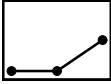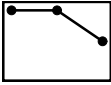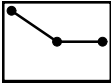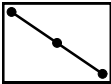
| profile | TP1 TP2 TP3 | number of genes with log2FC >1 Bienertia | number of genes with log2FC >1 Suaeda |
|---------|-------------|------------------------------------------|---------------------------------------|
| 1 | | 178 | 287 |
| 2 | | 880 | 647 |
| 3 | | 58 | 49 |
| 4 | | 613 | 854 |
| 5 | | 375 | 962 |
| 6 | | 20 | 26 |
| 7 | | 430 | 710 |
| 8 | | 77 | 286 |
| DE genes | | 2631 | 3821 |

Table 3.5: Overlap between up- and downregulated genes in Bienertia and Suaeda. DE = differentially expressed

| | up in Suaeda | down in Suaeda |
|---|---|---|
| upregulated in Bienertia | 29 % | 7 % |
| downregulated in Bienertia | 2 % | 20 % |
| not DE in Bienertia | 69 % | 73 % |

| | up in Bienertia | down in Bienertia |
|---|---|---|
| upregulated in Suaeda | 42 % | 5 % |
| downregulated in Suaeda | 12 % | 54 % |
| not DE in Suaeda | 46 % | 41 % |

### 3.4.3 Enrichment of MapMan BINs in differentially expressed genes

More than half of the differentially expressed genes are regulated differently in the two species. To obtain insight into which developmental processes are responsible for the difference between Bienertia and Suaeda, the differentially regulated genes were examined further. For a general overview of gene function in up- and downregulated genes, all differentially expressed genes with a log2 foldchange > 1 in Bienertia and Suaeda were sorted into primary BINs. For this, the profiles 1, 2, 4 and 5, 7, 8 were combined to represent the upregulated and downregulated genes, respectively. Figure 3.12 shows the number of genes assigned to the 15 most common primary BINs for each of the four situations (upregulated in Bienertia/upregulated in Suaeda/downregulated in Bienertia/downregulated in Suaeda). For most BINs among the upregulated genes there is little difference between both species, which could suggest functional synchronicity despite the high number of differently regulated genes. Genes assigned to the BINs hormone metabolism and protein are higher among the upregulated genes in Bienertia compared to Suaeda, while there are twice as many upregulated genes assigned to the BIN RNA in Suaeda compared to Bienertia. Unexpectedly, there are only a few genes associated with the BIN cell organisation, which comprises cytoskeletal components and regulators, differentially expressed in both species.

**Photosynthetic genes are upregulated early in Bienertia, but not in Suaeda**

Interestingly, a large number of genes assigned to the BIN photosynthesis are upregulated in Bienertia, but not Suaeda. At the same time, a higher number of photosynthetic genes are downregulated in Suaeda, but not Bienertia. The difference is still clearly noticeable after accounting for the twice as high number of downregulated genes in Suaeda. Previous studies have shown that some photosynthetic genes are expressed in etioplasts, while the expression of others is light-dependent (von Zychlinski et al. 2005; Kanervo et al. 2008). A cross-check with the light-induced photosynthetic genes in Arabidopsis shows that 50 of the 87 photosynthetic genes that are upregulated early in the dark-grown Bienertia seedlings are light-induced in Arabidopsis. 19 of these genes are upregulated after exposure to light in Bienertia.
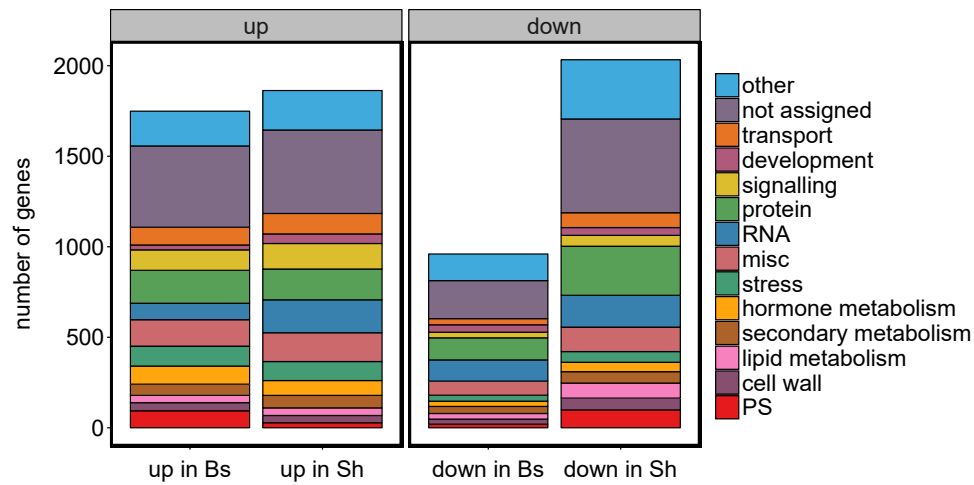
Figure 3.12: Comparison of annotated gene function in differentially expressed genes in Bienertia and Suaeda by MapMan primary BINs. Differentially expressed genes from both species were sorted into up- and downregulated. The colors indicate the associated BIN and the height of the colored blocks reflects the number of genes sorted into the BIN. Bs = Bienertia, Sh = Suaeda.

## $C_4$ cycle and related genes are upregulated early in dark-grown Bienertia cotyledons

Even though the cotyledons were undergoing skotomorphogenesis during the experiment, meaning the cotyledons contain non-photosynthetic etioplasts, not chloroplasts, several photosynthesis-related genes are upregulated early in Bienertia. While *B. vulgaris* does not perform $C_4$ photosynthesis, the $C_4$ cycle genes are nevertheless present in the genome. It has been shown that the $C_4$ gene homologs in $C_3$ species fulfill different functions and are recruited for the $C_4$ pathway in $C_4$ plants (Brown et al. 2011). To see whether $C_4$ genes are among early upregulated photosynthetic genes, the expression pattern and relative abundance of the core NAD-ME-type $C_4$ genes, $C_4$-related transport genes and several $C_3$ genes were compared between Bienertia and Suaeda (Table 3.6, Figure 3.13). The mean expression across the first three time points for both species was used to calculate the ratio between Bienertia and Suaeda and is from here on referred to as relative transcript ratio.

Both *PPDK* and *NAD-ME* are highly expressed during early development in Bienertia,

Table 3.6: Relative transcript ratio of $C_4$ and selected $C_3$ genes in Bienertia and Suaeda. The ratio was formed by dividing the mean expression of TP1 to 3 in Bienertia by the mean expression of TP1 to 3 in Suaeda.

| Gene | Name | relative transcript ratio |
| --- | --- | --- |
| Bv1_013550_fjqs.t1 | PPDK | 9.42 |
| Bv9_209750_xeaz.t1 | PEPC | 1.45 |
| Bv1_005810_wmrj.t1 | NAD-ME | 7.83 |
| Bv4_085990_rzto.t1 | NAD-MDH | 2.31 |
| Bv9_209080_aeyz.t1 | mAsp-AT | 10.73 |
| Bv8_196670_ezcs.t1 | cAsp-AT | 11.08 |
| Bv8_194450_rkme.t1 | ß-CA 2.1 | 11.00 |
| Bv6_148840_uffy.t1 | ß-CA 2.2 | 1.04 |
| Bv8_195530_sxjq.t1 | BASS | 7.15 |
| Bv8_187290_jqid.t1 | PPDK-RP | 3.31 |
| Bv7_169130_kwer.t1 | Ala-AT | 0.98 |
| Bv2_026820_qkgn.t1 | RbcS | 10.67 |
| Bv4_094590_jftx.t1 | GAPDH-UE1 | 3.08 |
| Bv1_014260_waoc.t1 | GAPDH-UE2 | 8.65 |
| Bv4_094290_jgpp.t1 | GOX | 0.38 |

but not in Suaeda, with a relative transcript ratio of 9.4 and 7.8, respectively (Table 3.6). Both genes are also upregulated in light (Figure 3.13). *PEPC* is equally and lowly expressed in both species during skotomorphogenesis and increases strongly after exposure to light in Bienertia. *NAD-MDH* expression decreases from TP1 to 3 in both species. Its mean expression is 2.3-fold higher in Bienertia and it is not induced by light. *PPDK-RP* is expressed 3.3-fold higher in Bienertia than in Suaeda, and follows a pattern similar to *PPDK*. However, the overall expression was lower (691/208 NC). Both the cytoplasmic and the mitochondrial *Asp-AT*, which are needed in the *NAD-ME* cycle to shuttle $C_4$-acids to the mitochondria, were much more abundant in the $SCC_4$ species than in Suaeda (10.7 and 11.1 fold, respectively) as well as overall highly expressed (20111/6686 NC). However, they are not upregulated in the light. *Ala-AT*, like *PEPC*, did not show an early upregulation in the dark, however, it was strongly upregulated in the light. One form of the ß-carbonic anhydrase (*ß-CA 2.1*) had a relative transcript ratio of 10.9, while also being highly expressed in Bienertia (2083 NC). The expression of the pyruvate-sodium symporter *BASS*, also associated with $C_4$ photosynthesis, was 7.1 times higher in Bienertia than in Suaeda. Both *BASS* and *ß-CA 2.1* follow the expression pattern of *PPDK* and *NAD-ME* and are upregulated after exposure to light. While the core $C_4$ genes are much more abundantly expressed in Bienertia than Suaeda, the same is true for several $C_3$ photosynthetic genes. The small subunit of *RuBisCO* and the *glyceraldehyde 3-phosphate dehydrogenase (GAPDH)*

show similar expression patterns. This is consistent with the earlier observation of upregulated genes in Bienertia, which show a different expression pattern for a large number of photosynthetic genes. In contrast, the photorespiratory gene glycolate oxidase (*GOX*) is more abundant in Suaeda. In Bienertia, several $C_3$ and $C_4$ genes are upregulated early during development in the dark. These transcripts are also more abundant in Bienertia compared to Suaeda. It is possible that the expression of photosynthetic genes occurs in anticipation of light exposure. While the transcriptional investment data shows an increase in photosynthetic expression at TP2 in Suaeda, the effect is less pronounced (Figure 3.11).
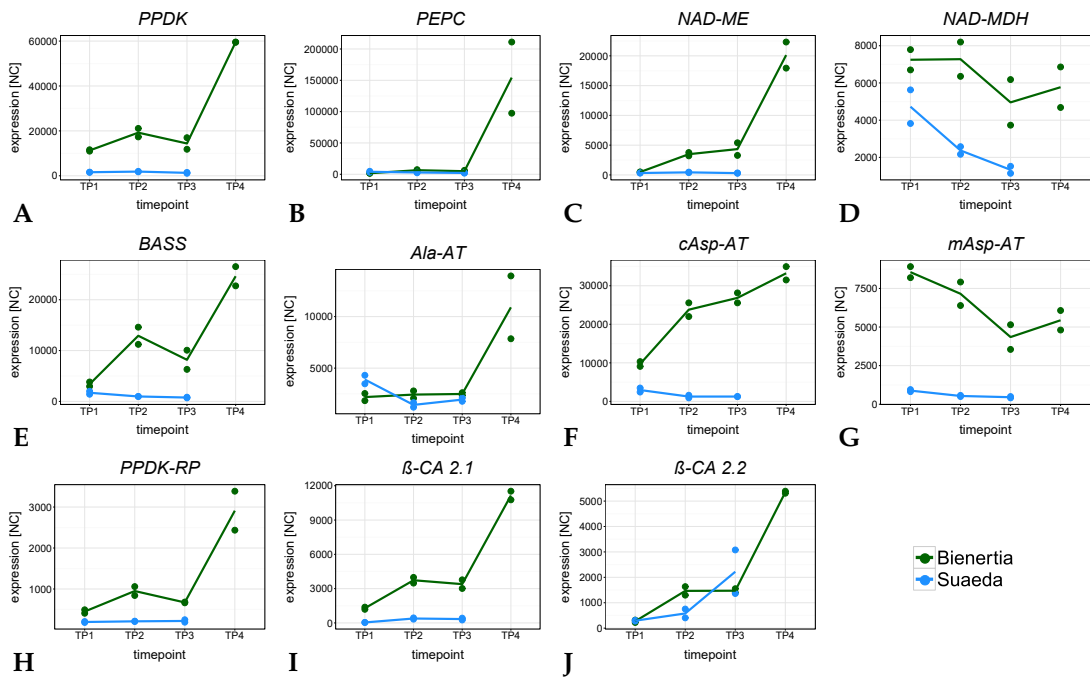


Figure 3.13: Expression of genes involved in $C_4$ photosynthesis during cotyledon development in Bienertia and Suaeda. The y-axis shows normalized counts as calculated by the DESeq2 package.

**Enrichment analysis of downregulated processes reveals differences in plastid processes**

Gene expression profiling showed that there is a much higher number in downregulated genes in Suaeda (Figure 3.12 and Table3.4). The downregulated genes were tested for enrichment with the PageMan module of the MapMan Software and com-

pared between both species (2.2.2). 45 BINs were enriched for downregulated genes in Bienertia (gene profiles 5, 7, and 8) (Section A.1). The primary BINs RNA and DNA make up a third of the enriched BINs. Development-related genes (storage proteins and late embryogenesis abundant) are also enriched among downregulated genes in Bienertia. In Suaeda, there are 111 enriched BINs for all downregulated genes (Section A.1). A closer look at the temporal course of downregulation reveals that most genes are downregulated between TP2 and 3 (gene profile 5) (Figure 3.14). Photosynthetic genes, plastid ribosomes, as well as genes involved in amino acid metabolism, tetrapyrrole and vitamin K synthesis are enriched in this profile. Notably, these processes are all at least partially localized to the plastid. In contrast, there is little downregulation of plastid-associated processes in Bienertia. It appears that these processes are regulated differently in Bienertia and Suaeda.
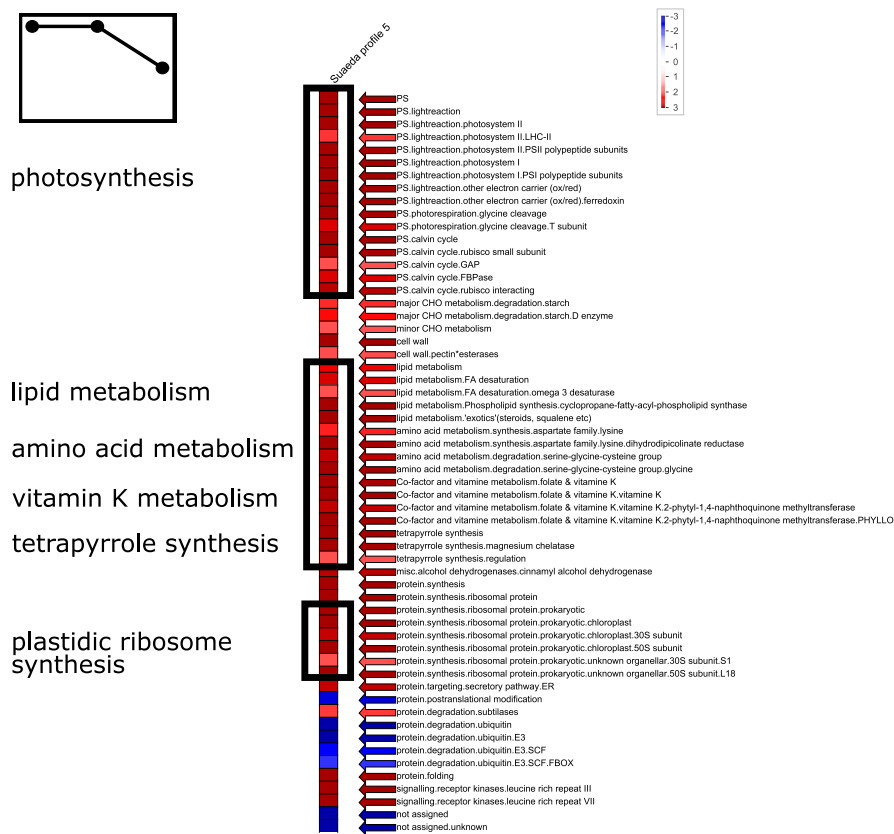


Figure 3.14: Enrichment analysis on genes downregulated late in Suaeda (Profile 5). The gene set was compared to all genes above the cutoff value using the PageMan module in MapMan. Red boxes show overrepresented BINs, blue boxes show underrepresented BINs.

## Plastidial ribosomes and tetrapyrrole synthesis

The expression of plastid ribosomes is a measure for the biosynthetical output of the plastid. To see whether the downregulation of plastid ribosomal genes (pRPs) is prevalent in older Suaeda cotyledons, the gene expression of 62 pRPs, comprising of both large (50S) and small (30S) subunit genes, was examined. For each gene, the expression was calculated relative to TP1. In Suaeda, the majority of pRPs peak at TP1 and are downregulated between TP2 and 3. A smaller group is first slightly upregulated, then downregulated to the expression level of TP1. Only a few genes are not downregulated from TP2 to 3. Overall, ribosome expression at TP2 and 3 is equal to or lower than TP1. In contrast, most pRPs in Bienertia, and specifically the group of downregulated pRPs in Suaeda, are upregulated from TP1 to TP2, then also show a downward trend to the approximate TP1 expression value. However, only six were identified as downregulated by the DESeq2 analysis. The genes showing an asynchronous pattern in Suaeda behave similarly in Bienertia. About 32 % of pRPs show a different expression pattern in both species. As ribosome expression is typically correlated with translation in plastids, this means that the plastid biosynthetical output over time decreases in Suaeda.

One of the main functions of etioplasts is the production of chlorophyll precursors, so that chlorophyll is produced rapidly upon illumination. Chlorophyll synthesis shares the common part of the tetrapyrrole pathway with other tetrapyrroles like heme and siroheme. The intermediate metabolite Mg-protoporphyrin IX, and several genes (HY1/GUN2, HY2/GUN3, and GUN5) have been implicated in retrograde signaling and shown to influence nuclear gene expression (Mochizuki et al. 2001; Strand et al. 2003; Surpin et al. 2002). Genes associated with tetrapyrrole synthesis are enriched in profile 5 in Suaeda. A closer look at the 39 genes involved in this pathway shows that nearly all chlorophyll synthesis genes, both regulatory genes and several genes of the common branch are downregulated in Suaeda (44 %). In contrast, only 4 genes (10 %) are downregulated in Bienertia. It is possible that these genes are downregulated in Suaeda, because etioplasts in older cotyledons contain high enough levels of chlorophyll precursors. Taken together, the distinct expression of plastid ribosomes and genes of the tetrapyrrole pathway could reflect a different rate of plastid development in Bienertia and Suaeda.

Table 3.7: Changes in expression for genes involved in tetrapyrrole synthesis in Bienertia and Suaeda. Genes identified as differentially expressed are marked with gray boxes. Genes and categorization by pathway are based on MapMan.

| | | log2 foldchange | | | |
| | | Bienertia | | Suaeda | |
| function | pathway | TP1-2 | TP2-3 | TP1-2 | TP2-3 |
|---|---|---|---|---|---|
| glutaminyl-tRNA synthase | base | 0.15 | -0.43 | -0.96 | -0.86 |
| Glu-tRNA aminotransferase subunit B, GTAB | base | 0.29 | -0.86 | -0.26 | -0.68 |
| Glutamyl-tRNA reductase 1, GluTR | base | 0.45 | 0.07 | 0.43 | 0.22 |
| Glutamate-semialdehyde-aminomutase, GSA | base | 0.61 | -0.62 | -0.53 | -1.33 |
| Delta-aminolevulinic acid dehydratase, ALADH | base | -0.01 | -0.84 | -0.86 | -0.94 |
| Porphobilinogen deaminase, PGB | base | 0.57 | -0.97 | -1.18 | -1.40 |
| uroporphyrinogen-III synthase | base | -0.24 | -0.48 | -0.21 | -0.23 |
| Uroporphyrinogen decarboxylase, UDP | base | 0.33 | -0.83 | -0.63 | -1.73 |
| uroporphyrinogen decarboxylase, HEME 1 | base | 0.02 | -0.71 | -1.32 | -1.72 |
| Coproporphyrinogen III oxidase | base | -0.41 | -0.80 | -0.93 | -1.47 |
| Radical SAM superfamily protein | base | -0.25 | -0.95 | -1.83 | -0.52 |
| protoporphyrinogen oxidase, PPOX | base | 0.38 | -0.52 | -0.51 | -0.92 |
| Protoporphyrinogen oxidase, POX2 | base | -0.04 | -0.34 | 0.18 | -1.02 |
| Protoporphyrinogen oxidase, POX2 | base | -0.30 | -0.35 | 0.31 | -1.20 |
| magnesium chelatase, GUN5 | chlorophyll | 1.49 | -0.98 | -0.13 | -2.53 |
| Mg-protoporphyrin IX chelatase, CHLD | chlorophyll | 0.75 | -0.57 | 0.34 | -1.07 |
| Mg-protoporphyrin IX chelatase, CHLI | chlorophyll | 0.57 | -0.54 | -0.87 | -1.41 |
| Mg-protoporphyrin IX methyltransferase, CHLM | chlorophyll | 0.67 | -0.62 | -0.25 | -1.90 |
| Mg-protoporphyrin IX monomethyl ester cyclase | chlorophyll | 0.69 | -0.67 | -0.15 | -1.91 |
| 3,8-divinyl protochlorophyllide reductase, PCB2 | chlorophyll | -0.10 | -1.18 | -0.47 | -1.64 |
| protochlorophyllide oxidoreductase A, PORA | chlorophyll | 0.36 | -1.43 | 0.69 | -0.91 |
| protochlorophyllide oxidoreductase A, PORA | chlorophyll | 0.17 | -0.66 | -0.08 | -1.37 |
| chlorophyll synthetase, CHLG | chlorophyll | 0.51 | -0.94 | -0.90 | -1.67 |
| chlorophyllide oxygenase, CHLORINA 1 | chlorophyll | 1.20 | -0.23 | 1.26 | -0.47 |
| genomes uncoupled 4, GUN4 | regulation | 0.56 | -0.47 | -0.06 | -1.74 |
| coiled-coil, TPR domain containing protein, FLU | regulation | 0.03 | -0.67 | -0.73 | -1.11 |
| Ferrochelatase-2, FC2 | heme | -1.12 | 0.46 | -0.23 | 0.26 |
| Ferrochelatase-2, FC2 | heme | -0.98 | 0.27 | -0.30 | 0.36 |
| Ferrochelatase-2, FC2 | heme | 0.92 | -0.18 | -0.10 | -1.13 |
| heme oxygenase 2, HO2 | heme | 0.18 | -1.20 | -0.90 | -1.43 |
| heme oxygenase, TED4 | heme | 0.92 | 0.07 | 0.73 | 0.55 |
| urophorphyrin III methylase | siroheme | -0.64 | 0.05 | -0.62 | 1.13 |
| sirohydrochlorin ferrochelatase B, SIRB | siroheme | 0.10 | 0.14 | 0.67 | 0.00 |
| sirohydrochlorin ferrochelatase B, SIRB | siroheme | 0.12 | 0.12 | 0.84 | -0.03 |
| Flavodoxin family protein | unspecified | -0.05 | -0.06 | 0.24 | 0.55 |
| SOUL heme-binding family protein | unspecified | 0.04 | -1.92 | -1.24 | -1.63 |
| cytochrome c oxidase 10, COX10 | unspecified | -0.49 | 0.52 | -0.35 | -0.21 |
| SOUL heme-binding family protein | unspecified | 0.32 | 0.08 | 0.54 | 0.35 |
| chlorophyllase 2, CLH2 | unspecified | NA | NA | 1.28 | 0.35 |

**A** Bienertia           **B** Suaeda

— other   — enriched in Suaeda profile 5   — asynchronous
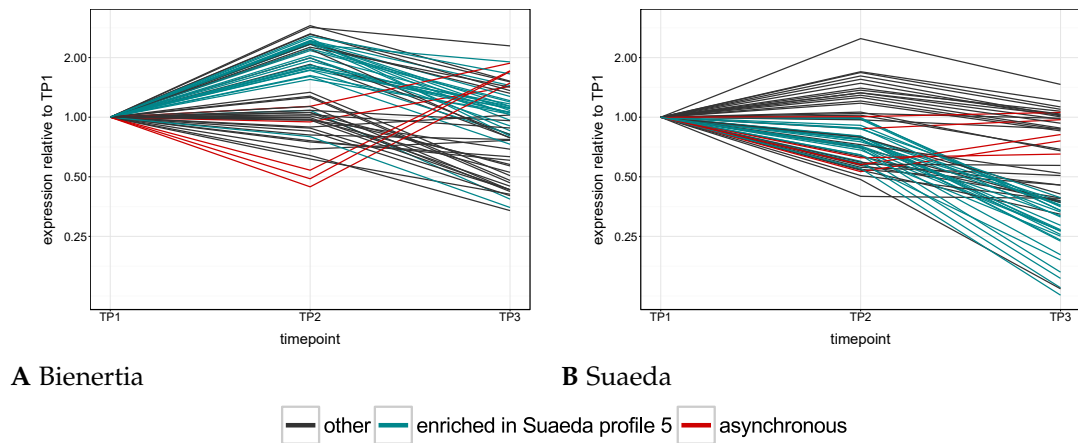
Figure 3.15: Gene expression of plastid ribosomal proteins during development in Bienertia (**A**) and Suaeda (**B**). The y-axis shows the expression relative to TP1 on a log-scale. Lines in blue represent the differentially expressed genes in Suaeda profile 5. Red lines represent the ribosomal proteins with an asynchronous expression pattern in both species. Black lines represent all other ribosomal proteins.

## 3.5 Identification of regulators and effectors of SCC$_4$ morphology

The aim of this thesis is to find the genes responsible for the SCC$_4$ morphology in the chlorenchyma cells of Bienertia. At TP2, a pre-CC is already visible in the chlorenchyma cells (Section 3.2), which means that the genes responsible are likely to be upregulated between TP1 and 2 and still expressed at TP3. The PCPs appear to move to their final positions between TP2 and 3. Accordingly, the expression for the genes involved in this process should increase in the later stages. However, genes with a regulatory function may be expressed earlier. A *k*-means clustering approach and filtering based on expression patterns of differentially expressed genes are employed to identify genes of interest and transcription factors. In addition, genes of interest specific to Bienertia are identified based on the read mapping to the Bienertia transcriptome. Finally, the genes more abundant in either species based on the average gene expression across TP1 and 3 are identified and can be found in an abbreviated form in Table A.1. The complete list of genes that show at least a fivefold difference in average expression in either species can be found in Table 5 of the electronic supplemental material.

### 3.5.1 *k*-means clustering

Clustering is a frequently employed strategy in the analysis of RNA-seq data to find general expression patterns, since co-expression of genes can point to a shared function or pathway (van Noort et al. 2003). *k*-means clustering was performed on the dataset of differentially expressed genes to find an expression profile that fit candidate criteria, namely genes upregulated in Bienertia, with unchanging or downregulated expression in Suaeda. This resulted in ten clusters that contain 133 to 314 genes. Cluster eight and ten (Figure 3.16) contain 225 and 195 genes, respectively, which show an increase in expression from Bienertia TP1 to 3 and low and unchanging expression in Suaeda. These genes could conceivably play a role in $SCC_4$ development, as gene expression is synchronous to the observed changes in morphology, while they do not seem vital to the development in the $C_3$ plant. The complete list of *k*-means clusters is available in Table 3 of the electronic supplemental material.
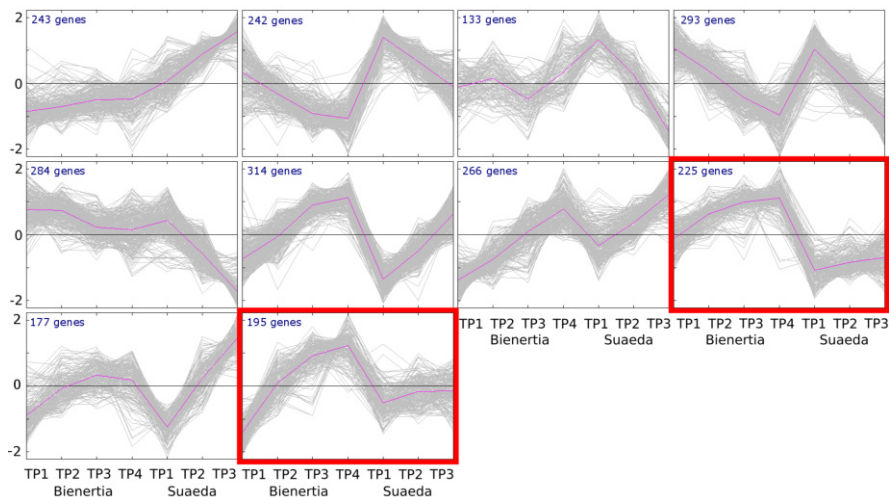


Figure 3.16: Cluster analysis of gene expression profiles in Bienertia and Suaeda. *k*-means clustering of all differentially expressed genes. The data was centered for each gene to allow for comparison between the species. Red box mark the clusters of interest for $SCC_4$ development.

### 3.5.2 qPCR

Several of the genes in clusters eight and ten were further tested by qPCR to confirm the general expression pattern. They were chosen based on high homology to both

the Bienertia transcriptome and the Suaeda *de novo* assembly, as well as containing a complete ORF. Species-specific primers were designed and qPCR was performed on all four biological replicates and 4 time points. For both species, the highly expressed EF2 was used to normalize the data so that a comparison across species was possible. Generally, the qPCR results confirmed the expression pattern found in the RNA-seq experiment (Figure 3.17). Surprisingly, the expression for *NAD-ME* is the same for both species, which also clearly indicates that it is light-induced. This is in contrast to the RNA-seq data. It is possible that in the case of Bienertia, a non-$C_4$ isoform was amplified.

### 3.5.3 Filtering criteria

To narrow down the number of genes identified by the k-mean clustering approach, an independent strategy was developed. Several filtering criteria were applied to all differentially expressed genes with a foldchange of at least two and an adjusted p-value of less than 0.1. Firstly, only genes upregulated at least two-fold in Bienertia between TP1 and 2, 2 and 3 or 1 and 3 were considered. Furthermore, these genes had to be not upregulated in Suaeda or generally higher expressed in Suaeda compared to Bienertia. Light-induced genes, i.e. those upregulated at least two-fold between TP3 and 4 were also not further considered. The remaining 220 genes were plotted individually and inspected visually which resulted in the removal of an additional 52 genes with similar expression patterns in Suaeda and Bienertia. Finally, genes with a CV higher than 40 % and low expression were not included in the final list. Based on these filter criteria 126 candidates were identified that could be involved in the formation of SCC$_4$ morphology. Of these, 120 were found in the aforementioned *k*-means clusters (Figure 3.16). They are listed in Table 3.9. Additional information on the candidate genes can be found in Table 4 of the electronic supplemental material.

### 3.5.4 Candidate list

The candidate list that was created using the aforementioned filter criteria contains 126 *B. vulgaris* genes. Five of those genes have isoforms that also matched the criteria. A BLAST search against the Bienertia and Suaeda transcriptome found 101 and 104 unique gene homologs, respectively. That means several of the *B. vulgaris* genes
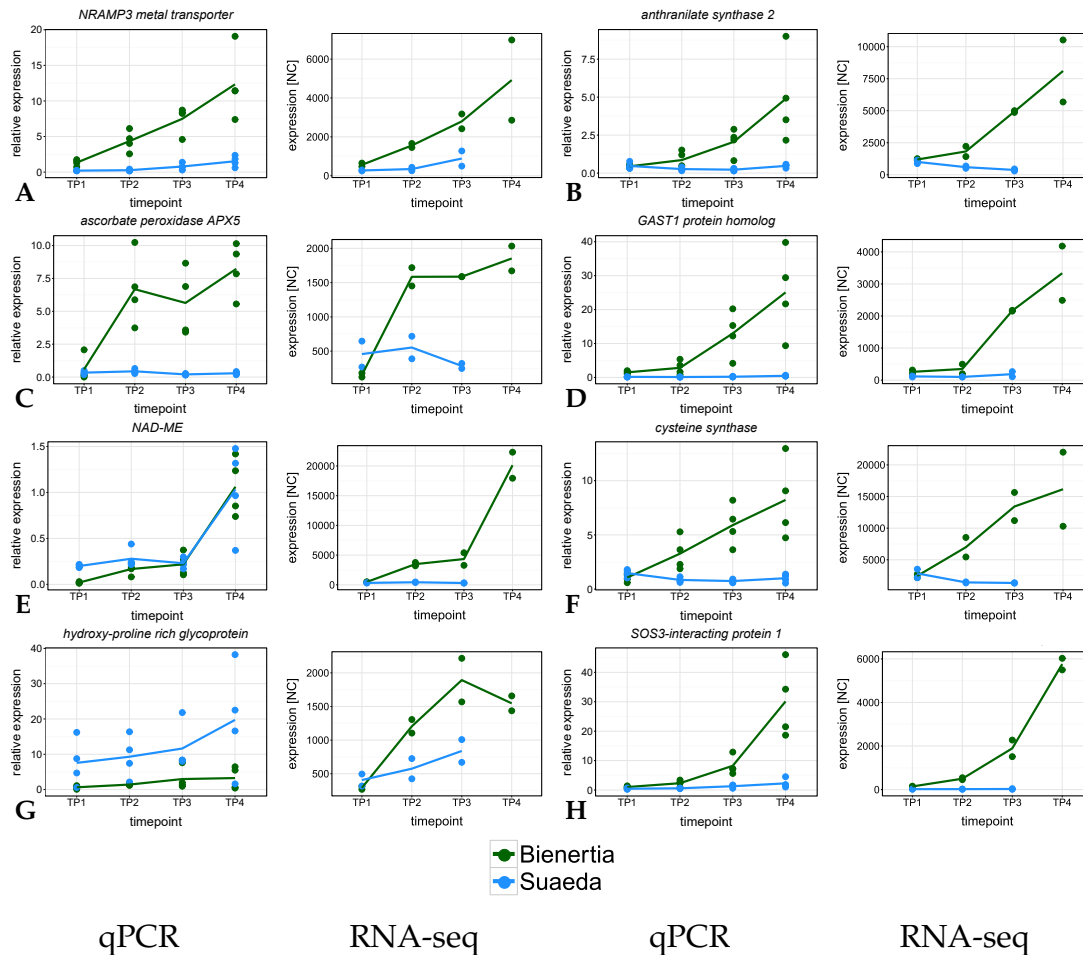
Figure 3.17: Comparison of RNA-seq and qPCR data for selected genes. qPCR data is shown in columns 1 and 3, RNA-seq data is shown in columns 2 and 4. qPCR was performed with species-specific primers on the respective gene homologs. All qPCR data was normalized by *elongation factor 2 (EF2)*. The y-axis for RNA-seq expression shows NC. **A**: *NRAMP metal transporter 3*. **B**: *anthranilate synthase*. **C**: *ascorbate peroxidase 5 (APX5)*. **D**: *GAST1 protein homolog* **E**: *NAD-malic enzyme (NAD-ME)*. **F**: *cysteine synthase*. **G**: *hydroxy-proline rich protein*. **H**: *SOS3-interacting protein 1.*

were homologous to only one Bienertia or Suaeda contig. For 14 of the candidate genes, a Bienertia homolog could not be identified. However, for the majority of genes, homologous Bienertia genes were found and at least partly verified by reverse BLAST. With the exception of one Bienertia contig, all identified homologs contained at least one ORF. The functional annotation of the candidates is illustrated in Figure 3.18. 33 genes were not assigned a MapMan BIN. A search for conserved motifs or domains was performed for these genes, but did not yield any results. There are nine predicted transcription factors and eight genes involved in signaling among the candidates. Four genes were annotated with a function in development or cell organization. Previous studies have shown that the cytoskeleton is essential for maintaining chloroplast position in mature cells, however, the details of the interaction between cytoskeleton and chloroplast are unknown. Among the candidate genes are two calmodulin-like proteins, which are involved in cytoskeleton organization and an IQ-domain containing protein which can bind to microtubules. For eight genes, associated GO-terms suggest a plastid localization. For four of them, TargetP confirms the localization, while predicting ten more genes to localize to the chloroplast, 13 to the mitochondria and 33 to the ER. 38 genes are significantly upregulated early, i.e. between TP1 and 2, however, only one is upregulated between TP2 and 3. The difference is likely due to the high variance of the TP3 samples, as this influences the calculation of differential expression. Among the early upregulated genes are two transcription factors, phytochrome-associated protein 2 (PAP2) and xylem NAC domain 1 (XND1). PAP2 is an auxin-regulated gene implicated in chloroplast differentiation (Listiawan et al. 2015). A chalcone synthase, which catalyzes a central step in flavonoid biosynthesis, is also upregulated early. Flavonoids fulfill many functions in plants, including pigmentation and mediation of growth (Besseau et al. 2007; Mol et al. 1998). It is possible that the upregulation of the chalcone synthase is related to the reddish cotyledon phenotype observed in Bienertia.

**Actin 7 is the isoform higher expressed in Bienertia**

An actin isoform was identified as possible effector of $SCC_4$ morphology, which contained only a partial ORF. It was annotated as actin 2 (ACT2) by the Mercator tool, however, a BLASTX search showed high similarities to other actin isoforms in other species. This is not surprising, as actin isoforms are highly conserved, especially on protein level (Šlajcherová et al. 2012). As the mapping to *B. vulgaris* was performed
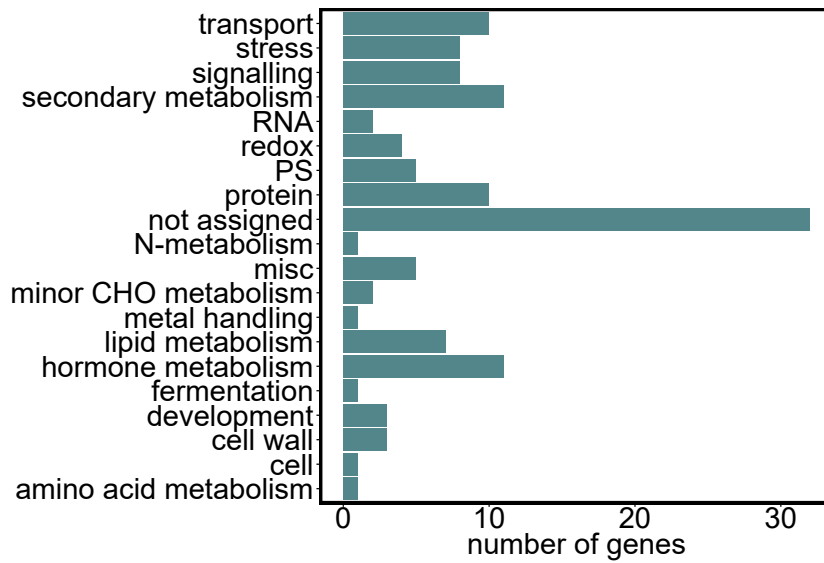
Figure 3.18: Functional annotation of candidate genes. Primary MapMan BINs were assigned to each candidate gene.

in protein space, this could mean that the expression of an individual isoform is inaccurate due to the reads matching other isoforms equally well. In order to identify the correct isoform, the nucleotide space read mapping to the Bienertia transcriptome was used (Section 3.3.1). Among the 15 actin sequences in Bienertia, two contain a complete ORF, while 13 are partial sequences. Four of these partial genes are highly expressed in Bienertia and show a similar gene expression pattern. In Bienertia, expression increases from TP1 to 3 approximately 1.5 to twofold. Alignment of the partial sequences showed that the overlapping parts of the sequence are identical on the nucleotide level (Figure 3.19). The partial contigs were assembled into a complete ORF. Taken together, it can be inferred that a single isoform of actin is highly expressed in Bienertia etiolated cotyledons. To compare levels of actin gene expression in Bienertia and Suaeda, which is not reliable in the CLC mapping due to the mapping bias, the NC of all actin sequences in *B. vulgaris* were accumulated by species. There is a twofold increase of actin expression in Bienertia compared to Suaeda (Table 3.8). The closest BLASTX match to the newly identified actin sequence from all four contigs is actin 7 (ACT7).
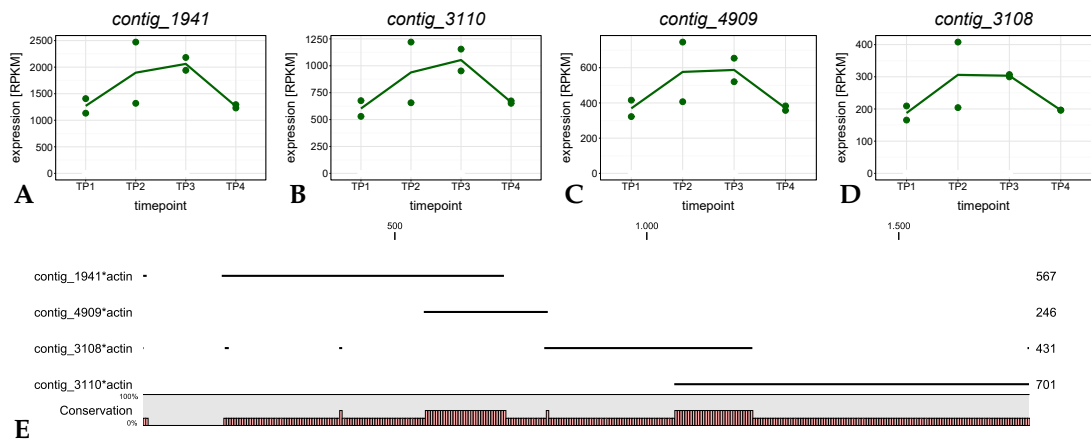
Figure 3.19: Assembly and expression pattern of a new actin isoform in Bienertia. **A-D**: Expression profiles of the four highly expressed actin genes in Bienertia. **F**: Alignment of highly expressed partial actin sequences in Bienertia. The red bars below the alignment indicate identical nucleotides between sequences.

Table 3.8: Average expression of *B. vulgaris* actin genes across TP1 to 3 in Bienertia and Suaeda.

| actin genes | average expression Bienertia | average expression Suaeda |
|---|---|---|
| Bv1_005380_uwpa.t1 | 3039.09 | 3450.40 |
| Bv1_005390_xmee.t1 | 78402.99 | 37034.46 |
| Bv2_024660_pyjn.t1 | 3427.25 | 1344.49 |
| Bv6_151450_cndk.t1 | 850.99 | 780.63 |
| Bv7_173970_ugxo.t1 | 3218.59 | 2324.56 |
| Bv_026560_mtjm.t1 | 67.86 | 7.58 |
| Bv_027930_ayhx.t1 | 1422.64 | 97.84 |
| Bv_034180_wpiy.t1 | 0.00 | 0.21 |
| Bv_035900_esqr.t1 | 67.50 | 3.07 |
| sum NC | 90496.92 | 45043.23 |
| ratio Bienertia/Suaeda | 2.01 | |
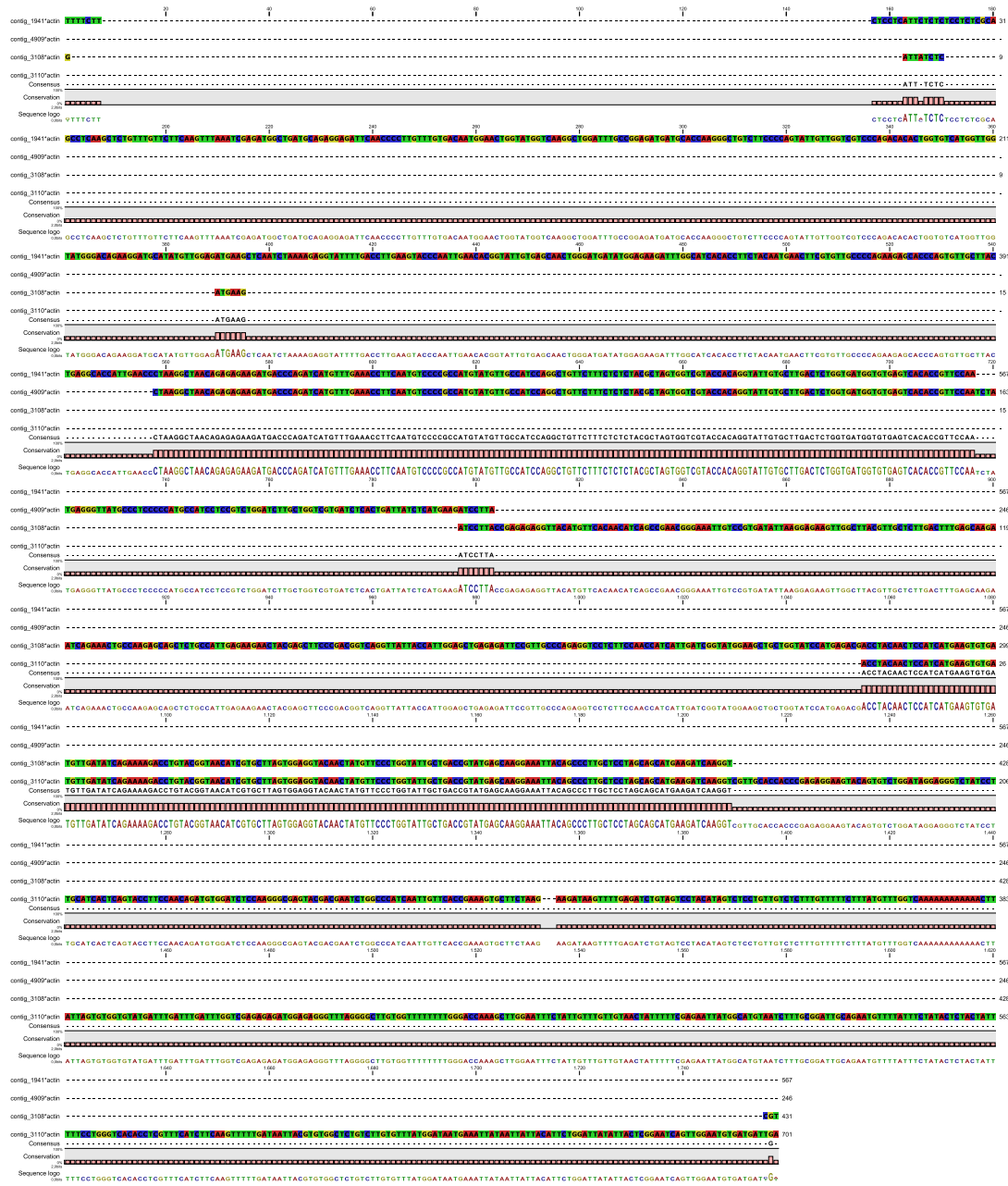
Figure 3.20: Full alignment of highly expressed partial actin sequences in Bienertia on nucleotide level.

Table 3.9: List of genes potentially involved in SCC$_4$ morphology development, specifically formation of the CC. The log2 foldchange refers to the estimated foldchange between TP1 and 3. The predicted function is the combined result of BLAST results, Blast2GO and Mercator annotation.

| contig | expression | log2FC | predicted function |
| --- | --- | --- | --- |
| Bv_026560_mtjm.t1 | medium | 8.04 | actin 7 (ACT7) |
| Bv_001120_qwkq.t1 | medium | 5.90 | auxin-binding protein ABP19a |
| Bv_011570_zmmm.t1 | medium | 5.64 | gibberellin 20-oxidase |
| Bv5_104770_htwy.t2 | medium | 5.48 | ent-kaurenoic acid hydroxylase |
| Bv5_104770_htwy.t1 | medium | 5.25 | ent-kaurenoic acid hydroxylase, gibberellin synthesis |
| Bv5_104780_kudm.t1 | high | 5.17 | ent-kaurenoic acid hydroxylase |
| Bv5_099020_yunk.t1 | medium | 5.07 | unknown |
| Bv_038060_jhdu.t1 | high | 4.98 | unknown |
| Bv_005070_jjst.t1 | high | 4.92 | UDP-Glycosyltransferase superfamily protein |
| Bv_016310_wuxy.t1 | high | 4.84 | unknown |
| Bv2_030050_fnzz.t1 | medium | 4.81 | isoflavone 2 -hydroxylase-like |
| Bv6_138820_zrmw.t1 | medium | 4.72 | carbon/nitrogen Insensitive1 (CNI1) |
| Bv5_099100_cywa.t1 | medium | 4.54 | unknown |
| Bv_038700_uxcg.t1 | high | 4.52 | unknown |
| Bv8_190180_ktgn.t1 | medium | 4.43 | xylem NAC domain 1 (XND1) |
| Bv6_132650_cqgk.t1 | medium | 4.22 | unknown |
| Bv2_030040_dnyj.t1 | medium | 4.10 | isoflavone 2 -hydroxylase-like |
| Bv2_041940_pjau.t1 | medium | 3.97 | sulfotransferase 2A |
| Bv3_058210_feuz.t1 | high | 3.88 | acidic endochitinase |
| Bv_042280_sjsm.t1 | high | 3.88 | unknown |
| Bv2_042500_ciwn.t1 | high | 3.79 | chalcone synthase 2 |
| Bv5_099060_ytgx.t1 | medium | 3.77 | unknown |
| Bv6_152180_ogah.t1 | medium | 3.74 | AMP-dependent synthetase and ligase family protein |
| Bv6_149030_cndy.t1 | medium | 3.69 | xyloglucan:xyloglucosyl transferase 33 (XTH33), cell wall modification |
| Bv2_042490_cfac.t1 | high | 3.65 | chalcone synthase |
| Bv6_152180_ogah.t2 | medium | 3.64 | AMP-dependent synthetase and ligase family protein |
| Bv2_042480_pyot.t1 | high | 3.64 | chalcone synthase 2 |
| Bv8_195280_gwfj.t1 | medium | 3.62 | unknown |
| Bv4_091130_ekaa.t1 | medium | 3.62 | chaperone protein dnaj chloroplastic-like |
| Bv_025990_iata.t1 | high | 3.58 | unknown |
| Bv8_197650_dtjf.t1 | high | 3.46 | unknown |
| Bv1_008690_mmma.t1 | high | 3.44 | EXORDIUM like 2 (EXL2) |
| Bv9_211270_fxis.t1 | medium | 3.41 | stress responsive a b barrel domain |
| Bv6_130360_dzya.t1 | high | 3.39 | microsomal ascorbate peroxidase APX5 |
| Bv1_015340_gfmt.t1 | medium | 3.36 | gibberellin-regulated family protein |
| Bv3_049140_chgz.t1 | medium | 3.36 | probable inorganic phosphate transporter 1-3 |
| Bv4_096830_sumq.t1 | high | 3.30 | unknown |
| Bv6_130330_fgat.t1 | medium | 3.28 | microsomal ascorbate peroxidase APX5 |
| Bv4_096840_ajge.t1 | high | 3.24 | unknown |
| Bv8_195980_nkip.t1 | medium | 3.24 | heavy metal transport detoxification superfamily protein |
| Bv4_074190_ctkt.t1 | high | 3.20 | unknown |
| Bv4_096850_htfc.t1 | high | 3.19 | unknown |
| Bv6_134800_ifmu.t1 | high | 3.15 | white-brown complex homolog protein 11 (WBC11 |

| contig | expression | log2FC | predicted function |
| --- | --- | --- | --- |
| Bv4_074220_juow.t1 | medium | 3.15 | unknown |
| Bv3_062340_hpte.t1 | medium | 3.14 | gibberellin-regulated GASA/GAST/Snakin family protein |
| Bv5_127070_crpg.t1 | medium | 3.14 | EF hand calcium-binding protein family |
| Bv3_052930_ukic.t1 | medium | 3.13 | unknown |
| Bv2_035770_jzij.t1 | high | 3.11 | cinnamic acid 4-hydroxylase |
| Bv2_027410_wqky.t1 | medium | 3.03 | unknown |
| Bv7_177350_yaxh.t1 | medium | 3.02 | major facilitator superfamily protein |
| Bv7_176320_ecaz.t1 | high | 3.00 | ECERIFERUM 4 (CER4), cutin development |
| Bv3_068800_ijrw.t1 | medium | 2.95 | dirigent protein 22-like |
| Bv4_089230_uzik.t1 | high | 2.95 | mannose-binding lectin superfamily protein |
| Bv5_104940_zody.t1 | medium | 2.94 | alpha/beta-Hydrolases superfamily protein |
| Bv5_099030_jxuj.t1 | medium | 2.93 | cupredoxin superfamily protein |
| Bv6_141320_gkux.t1 | medium | 2.90 | anthocyanidin 3-o-glucosyltransferase 2-like |
| Bv_015720_eoke.t1 | high | 2.87 | unknown |
| Bv_018260_zyau.t1 | medium | 2.85 | ( )-reticuline 7-o-methyltransferase-like |
| Bv2_035420_ytmj.t1 | high | 2.83 | major facilitator superfamily protein |
| Bv7_157870_rssd.t1 | high | 2.83 | nuclear transport factor 2 (NTF2) family protein |
| Bv5_100780_rgrk.t1 | high | 2.83 | zeamatin |
| Bv5_100760_woqm.t1 | high | 2.82 | protein P21 |
| Bv7_169390_ykhy.t1 | medium | 2.82 | unknown |
| Bv6_149940_mjpc.t1 | medium | 2.81 | lipopolysaccharide-induced tumor necrosis factor-alpha factor homolog |
| Bv1_008720_jnpu.t1 | high | 2.78 | EXORDIUM |
| Bv5_100810_ufoi.t1 | high | 2.77 | protein P21 |
| Bv8_201590_duyy.t1 | high | 2.73 | quinone reductase family protein |
| Bv6_130400_nqyz.t1 | medium | 2.73 | microsomal ascorbate peroxidase APX5 |
| Bv3_055000_uksh.t1 | high | 2.70 | hydroxyproline-rich glycoprotein family protein |
| Bv3_068820_ipua.t1 | medium | 2.69 | dirigent protein 22-like |
| Bv7_177350_yaxh.t2 | medium | 2.69 | major facilitator superfamily protein |
| Bv4_075220_usks.t1 | high | 2.68 | photosystem I reaction center PSI-L |
| Bv4_074120_dmgk.t1 | medium | 2.63 | galactose oxidase/kelch repeat superfamily protein |
| Bv3_056180_nenr.t1 | high | 2.63 | glutamate dehydrogenase B |
| Bv3_067580_ukqh.t1 | medium | 2.62 | unknown |
| Bv2_047690_afok.t1 | medium | 2.58 | zinc finger (C3HC4-type RING finger) family protein |
| Bv9_219780_srym.t1 | high | 2.55 | isoflavone-7-O-methyltransferase 8 |
| Bv5_101230_ugwa.t1 | medium | 2.54 | homeobox protein HD1 |
| Bv2_029680_jmxw.t1 | medium | 2.53 | putative amino acid transporter |
| Bv8_182200_tgmq.t1 | medium | 2.50 | GDSL-like Lipase/Acylhydrolase superfamily protein |
| Bv2_029700_pfgu.t1 | medium | 2.49 | putative amino acid transporter |
| Bv2_034500_esgc.t1 | medium | 2.47 | glutaredoxin-c6-like |
| Bv4_082400_qdff.t1 | high | 2.45 | cysteine synthase |
| Bv4_092600_kenz.t1 | high | 2.43 | quinone reductase family protein |
| Bv5_100240_fyeq.t1 | high | 2.42 | brassinosteroid-responsive RING-H2 (BRH1) |
| Bv5_099040_fmtg.t1 | medium | 2.41 | unknown |
| Bv_001000_mguw.t1 | medium | 2.38 | 12-oxophytodienoate reductase 3, jasmonate biosynthesis |
| Bv8_192120_gcrt.t1 | high | 2.36 | ECERIFERUM 3 (CER3), cutin development |

Table 3.9: (continued)

| contig | expression | log2FC | predicted function |
|---|---|---|---|
| Bv9_223090_jjqj.t1 | high | 2.35 | cinnamoyl CoA reductase |
| Bv8_194380_dfuo.t1 | medium | 2.32 | unknown |
| Bv_028010_dyzz.t1 | medium | 2.32 | methyl esterase 2 |
| Bv2_029260_fpsj.t1 | high | 2.30 | photosystem II 5 kDa protein (PSII-T) |
| Bv2_027860_mpzo.t1 | medium | 2.30 | probable calcium-binding protein (CML24) |
| Bv6_135330_ctua.t1 | high | 2.29 | haloacid dehalogenase-like hydrolase (HAD) superfamily protein |
| Bv9_207100_waxj.t1 | high | 2.23 | sucrose non-fermenting-1-related protein kinase (subunit) |
| Bv9_219400_jqpq.t2 | high | 2.20 | major facilitator superfamily protein |
| Bv5_112010_hzyz.t1 | high | 2.19 | unknown |
| Bv1_004020_hacp.t1 | medium | 2.19 | WRKY Transcription Factor 22 |
| Bv5_106400_okjw.t1 | medium | 2.19 | calcium-dependent lipid-binding plant phosphoribosyltransferase family protein |
| Bv3_067310_hfde.t1 | medium | 2.18 | carboxyesterase 20 (CXE20) |
| Bv2_024290_rmpf.t1 | high | 2.17 | chloroplastic sedoheptulose-1,7-bisphosphatase |
| Bv7_166350_scxi.t1 | medium | 2.16 | pfkB-like carbohydrate kinase family protein |
| Bv6_143110_pgda.t1 | high | 2.15 | NAD(P)H dehydrogenase 18 (NDH18) |
| Bv8_196860_akgk.t1 | high | 2.15 | phytochrome-associated protein 2 |
| Bv6_130040_gkpy.t1 | medium | 2.14 | 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein |
| Bv2_025530_okoh.t1 | high | 2.14 | unknown |
| Bv4_076220_ewfn.t1 | high | 2.13 | Clade E-Growth-Regulating protein phosphatase 2C 2 (EGR2) |
| Bv5_098240_utms.t1 | medium | 2.09 | flavonoid 3 -monooxygenase-like |
| Bv3_070720_qekf.t1 | high | 2.08 | chloroplastic transketolase |
| Bv5_097770_utnr.t1 | high | 2.05 | nitrate transporter NRT1.1 |
| Bv9_219400_jqpq.t1 | medium | 2.03 | major facilitator superfamily protein |
| Bv7_166660_uiiq.t1 | high | 2.00 | leucine-rich repeat protein kinase family protein |
| Bv9_219650_uofe.t1 | medium | 1.97 | unknown |
| Bv7_175510_njtw.t1 | medium | 1.96 | F-box/LRR-repeat MAX2 homolog |
| Bv5_098250_rhah.t1 | medium | 1.96 | flavonoid 3 -monooxygenase-like |
| Bv2_025180_zzst.t1 | high | 1.93 | zinc finger (C3HC4-type RING finger) family protein |
| Bv3_055740_ueqe.t1 | high | 1.93 | chloroplastic transketolase |
| Bv4_086840_rres.t2 | medium | 1.92 | protein phosphatase 2C family protein |
| Bv4_086840_rres.t1 | medium | 1.90 | protein phosphatase 2C family protein |
| Bv4_086840_rres.t3 | high | 1.88 | protein phosphatase 2C family protein |
| Bv5_116750_zqhy.t1 | high | 1.85 | aldehyde dehydrogenase 2B4 |
| Bv2_036470_uude.t1 | high | 1.85 | IQ-domain 2 (IQD2) |
| Bv3_060910_hhoz.t1 | high | 1.84 | pectin lyase-like superfamily protein |
| Bv1_001460_srtz.t1 | high | 1.84 | C2 calcium/lipid-binding plant phosphoribosyltransferase family protein |
| Bv9_221230_dkzm.t1 | high | 1.84 | delta tonoplast intrinsic protein |
| Bv4_079230_nedr.t1 | high | 1.76 | pectinacetylesterase family protein |

### 3.5.5 Transcription factors

Transcription factors play a crucial role in development and are thus relevant to the central question of how $SCC_4$ morphology develops. Additionally, they are among the group of genes differentially regulated in both species, with a high number upregulated in Suaeda compared to Bienertia (Section 3.4.3). A closer look at the overall expression of transcription factors in the two species was thus warranted. The web-based tool PlantTFcat (Dai et al. 2013) was used to predict transcription factors in the *B. vulgaris* transcriptome. 3011 genes were annotated as transcription factors, transcriptional regulators or chromatin regulators. Of these, 1897 in Bienertia and 2023 in Suaeda, show an expression higher than 20 NC. The minority of the transcription factors is differentially regulated over time in both species (697 in Suaeda and 272 in Bienertia). There are proportionally more transcription factors upregulated in Suaeda than in Bienertia (470 and 170 genes, respectively), which confirms the results from the MapMan analysis (Section 3.4.3). There are less downregulated transcription factors in both species with 102 genes in Bienertia and 227 in Suaeda. If transcription factors are involved in the development of the $SCC_4$ morphology, they would likely be upregulated early in Bienertia, but remain lowly expressed or unchanged in Suaeda. In addition to the transcription factors identified in this manner, transcription factors constitutively expressed higher in Bienertia than in Suaeda are also considered to be of interest. The 56 transcription factors that fulfil these criteria are listed in Table 3.10. All transcription factors identified in *B. vulgaris* with a mean expression higher than 20 NC in either species are listed in Table 6 of the electronic supplemental material. They belong to 22 different transcription factor families, all of which have diverse roles in plant development. There are 11 genes belonging to the C2H2-type zinc finger transcription factor family. It is one of the largest transcription factor families in plants with 176 genes identified in Arabidopsis (Ciftci-Yilmaz et al. 2008) and involved in developmental processes, such as flower development, root hair formation and seedling development, as well as stress responses (Yun et al. 2002; Yan et al. 2014; Prigge et al. 2001; Tian et al. 2010; Liu et al. 2015). Basic helix-loop-helix (bHLH) transcription factors control various plant growth and developmental processes. Of the four bHLH transcription factors identified here, IBH1 and KIDARI both seem to be involved in regulating cell elongation. IBH1 has been shown to negatively regulate cell elongation (Ikeda et al. 2012; Zhiponova et al. 2014), while KIDARI (also named PRE1) has been identified as an antagonist to IBH1 in rice and Arabidopsis (Ikeda et al. 2012; Zhang et al. 2009). KIDARI is further involved in the repression

of light signaling in Arabidopsis (Hyun et al. 2006). Nine of the identified transcription factors belong to the Hap3/NF-YB transcription factor family, which is a small transcription factor family with only 10 members identified in Arabidopsis (Edwards et al. 1998) and 11 in rice (Thirumurugan et al. 2008). Hap3 proteins bind to CCAAT sequences in a promotor to control gene expression and also contain a histone-fold domain (Nardini et al. 2013). The best studied Hap3 gene is LEAFY COTYLEDON1 (LEC1) in Arabidopsis, which has been shown to control embryogenesis and is required for establishing cotyledon identity (Lotan et al. 1998; West et al. 1994; Kwong et al. 2003). In rice, OsHap3 was shown to regulate chloroplast biogenesis (Miyoshi et al. 2003). The Hap3 genes identified here are mostly classified as histone-like. The putative function of some of the identified transcription factors and their integration into $SCC_4$ development is discussed in section 4.1.3.

Table 3.10: List of transcription factors (TFs) potentially involved in $SCC_4$ development. The transcription factors listed were identified with three different methods. Nine transcription factors were identified as candidate genes, 20 were identified as differentially upregulated in Bienertia and not in Suaeda in addition to being at least twice as abundant in Bienertia. 27 were at least five times more abundant in Bienertia compared to Suaeda. Duplicates were removed. The ratio column refers to the ratio of average expression across TP1 to 3 between Bienertia and Suaeda.

| name | TF family | annotation | ratio |
|------|-----------|------------|-------|
| Bv8_196860_akgk.t1 | AUX-IAA | phytochrome-associated protein 2 (PAP2) | 24.97 |
| Bv5_100240_fyeq.t1 | C2H2 | brassinosteroid-responsive RING-H2 (BRH1) | 1.47 |
| Bv6_138820_zrmw.t1 | C2H2 | carbon/nitrogen insensitive 1 (CNI1) | 1.99 |
| Bv7_166350_scxi.t1 | C2H2 | pfkB-like carbohydrate kinase family protein | 1.39 |
| Bv2_025180_zzst.t1 | C2H2 | zinc finger (C3HC4-type RING finger | 1.46 |
| Bv5_101230_ugwa.t1 | HD-TALE-KNOX | Homeobox protein HD1 | 1.53 |
| Bv8_190180_ktgn.t1 | NAM | xylem NAC domain 1 (XND1) | 5.50 |
| Bv2_047690_afok.t1 | PHD | Zinc finger (C3HC4-type RING finger) family protein | 91.18 |
| Bv1_004020_hacp.t1 | WRKY | WRKY22 | 5.76 |
| Bv3_049290_kacx.t1 | AP2-EREBP | ERF/AP2 transcription factor family | 2.97 |
| Bv_001550_hzdz.t1 | AUX-IAA | indole-3-acetic acid inducible 9 (IAA9) | 2.11 |
| Bv8_189720_opmr.t1 | BED-type(Zn) | DAYSLEEPER | 2.11 |
| Bv5_118480_wtsa.t1 | bHLH | ILI1 binding bHLH 1 (IBH1) | 6.73 |
| Bv2_046530_kcci.t1 | bHLH | KIDARI (KDR) | 3.43 |
| Bv1_021380_gmre.t1 | bZIP | TGA1A-related gene 3 (TGA3) | 2.78 |
| Bv8_198850_yjyh.t1 | C2C2-CO-like | CONSTANS-like 5 (COL5) | 2.23 |
| Bv6_133170_cewx.t1 | C2H2 | RING/U-box superfamily protein | 2.11 |
| Bv_009860_qxar.t1 | C2H2 | NYC1-like (NOL) | 2.02 |
| Bv1_004250_kmdh.t1 | C2H2 | brassinosteroid-responsive RING-H2 (BRH1) | 2.01 |
| Bv8_201220_hxfs.t1 | GAGA-Binding-like | basic pentacysteine1 (BPC1) | 3.30 |
| Bv5_121400_txsu.t1 | Hap3/NF-YB | TBP-associated factor 4 (TAF4) | 2.81 |
| Bv5_121400_txsu.t2 | Hap3/NF-YB | TBP-associated factor 4 (TAF4) | 2.62 |
| Bv7_171150_hihf.t1 | HD-TALE-BEL | BEL1-like homeodomain 1 (BLH1) | 2.10 |
| Bv3_066350_xfje.t1 | HSF-type | heat shock transcription factor B2A (HSFB2A) | 3.18 |

Table 3.10: (continued)

| name | TF family | annotation | ratio |
|---|---|---|---|
| Bv3_060630_hyww.t1 | MYB-HB-like | telomere repeat binding factor 1 (TRB1) | 2.88 |
| Bv5_106480_gzwx.t1 | NAM | NAC domain containing protein 71 (NAC071) | 2.34 |
| Bv5_102730_momm.t1 | PHD | alfin-like 1 (AL1) | 2.39 |
| Bv2_047510_icgt.t1 | WD40-like | WD40 repeat-like superfamily protein | 3.14 |
| Bv8_200770_tois.t1 | WRKY | WRKY DNA-binding protein 35 (WRKY35) | 2.07 |
| Bv2_023470_huft.t1 | AP2-EREBP | ABA INSENSITIVE 4 (ABI4) | 6.88 |
| Bv9_220390_mmps.t1 | AUX-IAA | indole-3-acetic acid inducible 19 (IAA19) | 8.02 |
| Bv5_113380_euwo.t1 | bHLH | INDUCER OF CBF EXPRESSION 1 (ICE1) | 5.14 |
| Bv7_171140_qwks.t1 | bHLH | LONESOME HIGHWAY (LHW) | 283.99 |
| Bv3_053880_mpkc.t1 | bZIP | basic region/leucine zipper motif 53 (BZIP53) | 23.39 |
| Bv7_177640_aesm.t1 | C2C2-GATA | ZIM-LIKE 2 (ZML2) | 5.51 |
| Bv_001810_zdxq.t1 | C2H2 | RING-type Zinc finger protein | 15.54 |
| Bv1_015820_oeuc.t1 | C2H2 | RING/U-box superfamily protein | 5.15 |
| Bv6_128840_qhip.t1 | C2H2 | Major facilitator superfamily protein | 7.92 |
| Bv7_163120_gqkz.t1 | C2H2 | RING/U-box superfamily protein | 627.05 |
| Bv_001660_ufqc.t2 | C3H | Zinc finger C-x8-C-x5-C-x3-H type family protein | 257.02 |
| Bv9_210070_mkjc.t2 | CCHC(Zn) | F-box/RNI-like superfamily protein | 8.74 |
| Bv6_145210_uijh.t1 | FAR | FAR1-related sequence 11 (FRS11) | 289.55 |
| Bv5_107360_srki.t1 | GAGA-Binding-like | basic pentacysteine 7 (BPC7) | 12.93 |
| Bv_023460_sjyc.t1 | Hap3/NF-YB | Histone superfamily protein | 15.50 |
| Bv4_075650_krda.t1 | Hap3/NF-YB | histone H2A 12 (HTA12) | 74.20 |
| Bv4_086350_pcmw.t1 | Hap3/NF-YB | Histone superfamily protein | 15.04 |
| Bv4_086410_dhew.t1 | Hap3/NF-YB | Histone superfamily protein | 6.95 |
| Bv4_086420_hqgp.t1 | Hap3/NF-YB | Histone superfamily protein | 15.18 |
| Bv7_174390_koiq.t1 | Hap3/NF-YB | histone H2A 12 (HTA12) | 5.92 |
| Bv9_209670_zfng.t1 | Hap3/NF-YB | Histone superfamily protein | 5.43 |
| Bv5_121340_cuxq.t1 | JmjC | PKDM7D | 341.84 |
| Bv1_020230_ssoy.t1 | MYB-HB-like | Homeodomain-like superfamily protein | 8.07 |
| Bv_006020_syfo.t1 | PHD | VARIANT IN METHYLATION 1 (VIM1) | 9.51 |
| Bv_010750_yozs.t1 | WD40-like | WD-40 repeat family protein | 140.88 |
| Bv4_092160_nehd.t1 | WD40-like | homolog of yeast autophagy 18 (ATG18) | NA |
| Bv4_092160_nehd.t2 | WD40-like | homolog of yeast autophagy 18 (ATG18) | 921.51 |

## 3.5.6 Candidate genes unique to Bienertia

Because of the limitations inherent of cross-species mapping and comparison, the candidate list cannot by definition include genes that occur only in Bienertia, since all of the reads are mapped to the *B. vulgaris* transcriptome (Section 3.3.1). However, a putative "SCC$_4$ factor" could potentially be specific to the SCC$_4$ species Bienertia. Thus, Bienertia-only genes that match the gene expression profile based on the BLAT mapping to the Bienertia transcriptome were further explored. This was done by

filtering the expression profile of Bienertia genes that did not have a homolog in either the Suaeda or the *B. vulgaris* transcriptome. The same filter criteria were used as with the cross-species mapping to *B. vulgaris*. This resulted in 16 sequences that matched the criteria (Table 3.11). Twelve sequences are predicted to contain ORFs, however, only four sequences contain at least one complete ORF. A BLASTX search of all ORFs found gene homologs for just three of the sequences. To exclude the possibility that the identified sequences were wrongly assembled during the construction of the transcriptome, they were compared to an older transcript database (454 pyrosequencing). Nine sequences were highly similar, which indicates that these sequences might contain heretofore undescribed genes. However, a scan for protein motifs and conserved domains did not reveal potential functions for these sequences. Due to the fact that most ORFs were incomplete at the N-terminus, prediction of the subcellular location (plastid, mitochondria, secretory pathway) was possible for only a few. Two were predicted to contain a signal peptide, one a mitochondrial target peptide and three to be located elsewhere.

Table 3.11: List of candidate genes unique to Bienertia. * = incomplete ORF at start, # = incomplete ORF at end.

| gene | gene length | ORF length | predicted location | predicted function |
| --- | --- | --- | --- | --- |
| contig_1131 | 350 | none | | |
| contig_13988 | 285 | none | | flavonol sulfotransferase-like |
| contig_150523 | 331 | 327*# | | |
| contig_165637 | 445 | 174; 165* | S | hypothetical protein |
| contig_192472 | 228 | 228*# | | |
| contig_28058 | 234 | 228* | | |
| contig_28821 | 1075 | 219 | other | |
| contig_29695 | 1113 | 240; 156* | M | |
| contig_30083 | 341 | 330* | | |
| contig_42904 | 686 | none | | |
| contig_43361 | 304 | 303*#; 156* | | basic secretory protein family |
| contig_498492 | 265 | 183* | | |
| contig_56104 | 4142 | 177; 195; 153 | S; other; other | |
| contig_57257 | 230 | 192* | | |
| contig_786683 | 253 | 201* | | |
| contig_92129 | 348 | none | | |

### 3.5.7 Localization of BsCHUP1 in Bienertia chlorenchyma cells

CHUP1 is one of the known genes involved in chloroplast movement. In Arabidopsis, knock-out mutants are deficient in light-dependent chloroplast movement and chloroplasts are shown clumping together at the bottom of M cells (Oikawa et al. 2003). To investigate whether CHUP1 is involved in chloroplast positioning in Bienertia, a truncated CHUP1 gene containing only the N-terminus and the coiled-coil region, was fused to GFP and transiently expressed in Bienertia chlorenchyma cells under the control of the 35S-Promoter (Figure 3.21). In chlorenchyma cells overexpressing the truncated CHUP1-GFP protein, a loss of position of the PCPs could be observed. Instead of being randomly distributed around the cell periphery, PCPs cluster around the CC. The fusion protein appears to localize to the chloroplast envelope, however, it can not be discerned whether it localizes to all chloroplasts or specifically to the PCPs. Overexpression of 35S-GFP does not disrupt PCP localization and shows the expected localization pattern in cytosol and nucleus.

Arabidopsis CHUP1 was shown to contain several functional domains. Bienertia and Arabidopsis CHUP1 sequences were aligned and *in silico* analysis of protein domains and functional sites in the Bienertia CHUP1 sequence was performed to examine whether the Bienertia gene contains the same functional domains (Figure 3.21). The hydrophobicity plot shows that the N-terminal end of Bienertia CHUP1 contains a short hydrophobic region (Expasy ProtScale). A coiled-coil domain is predicted in the N-terminal part of the sequence (COILS, default settings). A protein domain scan with the Prosite Scan tool revealed the presence of two leucine zippers and a proline-rich region in the C-terminal part of the protein. The actin binding domain that was identified in Arabidopsis was not detected in Bienertia. However, the amino acid sequence of the proposed actin binding domain aligns perfectly with the Arabidopsis amino acid sequence, thus Bienertia likely also contains an actin binding site. Overall, the domains identified in Arabidopsis CHUP1 are also present in Bienertia CHUP1.

### Expression of CHUP1 and KAC1 and 2 in dark grown cotyledons

In dark-grown cotyledons of Bienertia and Suaeda, CHUP1 and KAC2, but not KAC1 are highly expressed. CHUP1 expression in both species is stable across TP1 to 3. After light exposure in Bienertia, CHUP1 expression increases considerably. KAC2 is expressed approximately three times higher in Suaeda than in Bienertia. Expres-
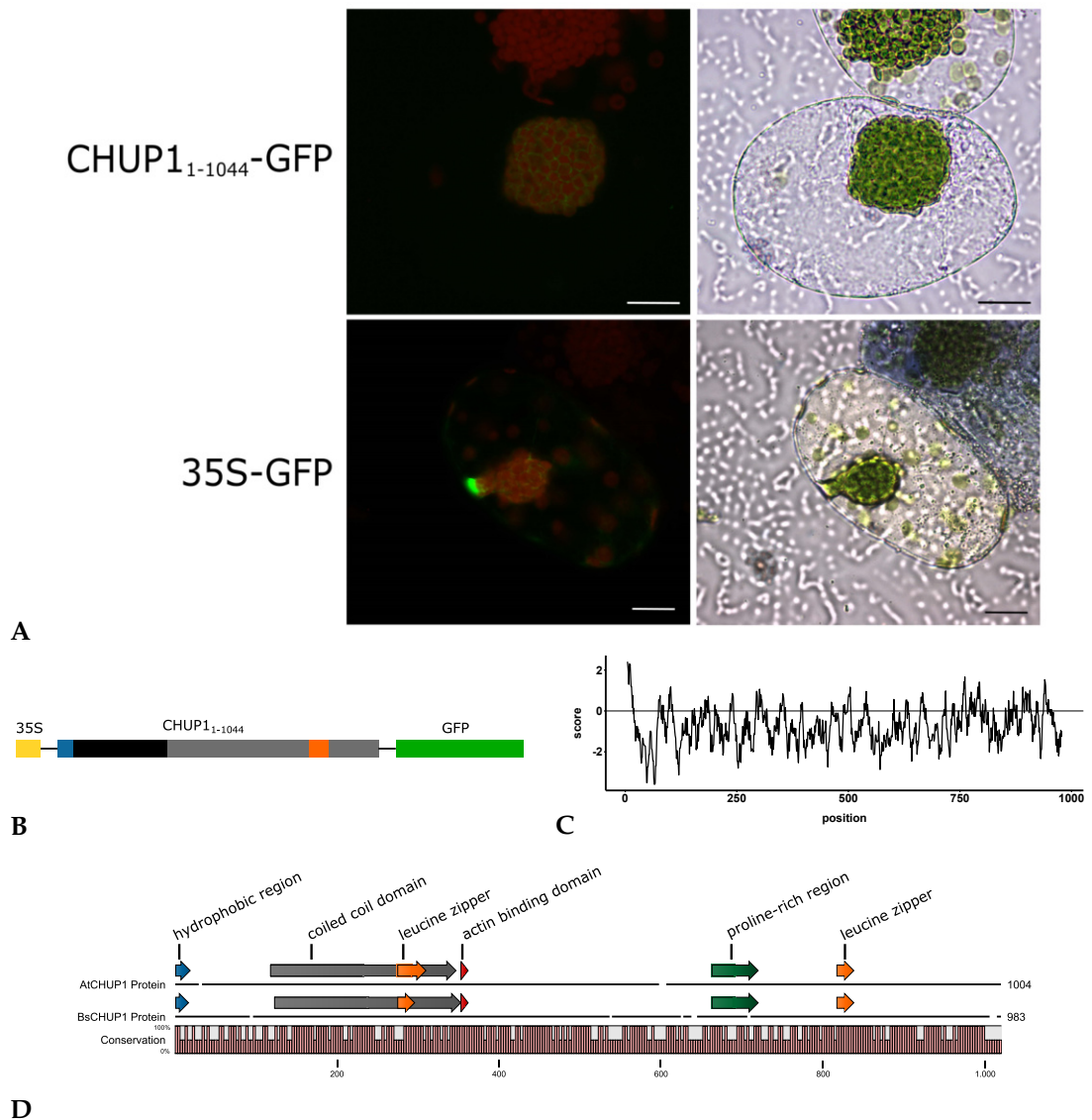
Figure 3.21: **A**: Localization of the truncated Bienertia 35S-CHUP1-GFP fusion protein. The first 1044 nucleotides were amplified, fused to GFP and transiently expressed under the 35S-Promotor (upper row, left). 35S-GFP was expressed as control (lower row, left). Scale bars = 10 µm. **B**: Schematic of the CHUP1$_{1-1044}$ fusion protein. The detected protein domains are indicated by colour. The hydrophobic N-terminal region is blue, the coiled-coil domain is grey and the leucine zipper domain is orange. **C**: Hydrophobicity plot of Bienertia CHUP1 protein sequence. The scores were calculated with the ProtScale tool using the Kyte-Doolittle scale with a window size of 9 and are shown on the y-axis. The x-axis shows the amino acid sequence. **D**: Alignment of Arabidopsis and Bienertia CHUP1 protein. The height of the light red bars at the bottom indicate the sequence consensus. The detected protein domains are indicated by colour. The annotations for Arabidopsis CHUP1 are adapted from Oikawa, 2003.

sion is stable across all time points in both species no light-induction can be observed in Bienertia. KAC1 is lowly expressed in Bienertia and Suaeda and does not increase expression after exposure to light. The high expression of CHUP1 and KAC2 in dark-grown cotyledons suggests that both genes might contribute to chloroplast positioning in the dark.
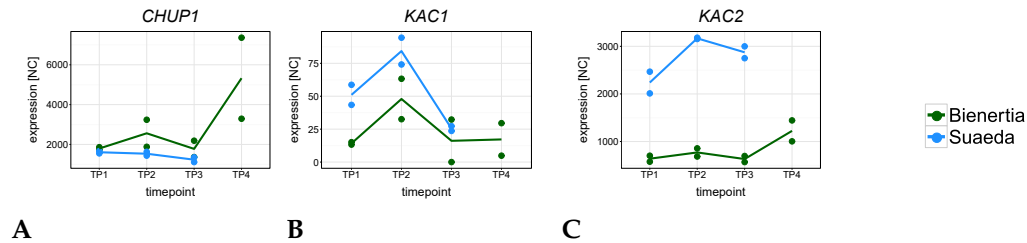


Figure 3.22: Gene expression patterns of known chloroplast movement genes in Bienertia and Suaeda. **A**: *CHUP1*, **B**: *KAC1*; **C**: *KAC2*. The y-axis shows NC.

# 4 Discussion

## 4.1 Identification of genes potentially involved in SCC$_4$ development

Chlorenchyma cells in dark-grown Bienertia cotyledons show development of the dimorphic compartments that are typical for chlorenchyma cells in mature photosynthetic tissue. In this thesis, a comparative transcriptomics experiment was performed that compares gene expression in three developmental stages of dark-grown cotyledons in Bienertia with cotyledon development in Suaeda, in order to find factors that regulate and effect SCC$_4$ development. Multiple genes showed changes in gene expression that reflect the morphological changes of SCC$_4$ development (Table 3.9). In addition, regulatory genes were identified by comparing overall abundance in both species (Table A.1). In the following paragraphs, some of those genes will be discussed in regard to their potential function in SCC$_4$ development.

### 4.1.1 Cytoskeleton-interacting factors are likely involved in chloroplast positioning

The cytoskeleton has been implicated in Bienertia chlorenchyma cell development (Park et al. 2009) and the maintenance of chloroplast positioning (Chuong et al. 2006; Park et al. 2009). The cytoskeleton in mature Bienertia cells shows a high degree of organization. Actin filaments radiate from the cell center to the periphery, while microtubules are arranged in a transverse pattern (Chuong et al. 2006; Park et al. 2009). In developing cells, actin strands associate with PCPs, while microtubules start to form cage-like structures around the CC (Park et al. 2009). However, it is unclear how exactly the cytoskeleton influences the initial chloroplast positioning process.

**Several regulators of microtubule organisation potentially contribute to CC formation**

Several genes of interest identified in this thesis are connected to the cytoskeleton (Table 3.9). They show increased expression in Bienertia compared to Suaeda and are upregulated during development of the CC. A homolog to the calmodulin-like protein (CML24) (Bv2_027860_mpzo.t1) was upregulated fivefold between TP1 and 3 in Bienertia (Table 3.9). In Arabidopis, CML24, a Ca2+-sensor, is reported to increase expression in response to a multitude of stimuli, including abscisic acid (ABA) treatment and mechanical stress (Delk et al. 2005). It has further been shown to alter microtubule orientation in root epidermal cells (Wang et al. 2011). Arabidopsis mutants lacking CML24 show a disorganized actin cytoskeleton in the pollen and pollen tube (Yang et al. 2014). A screen for cytoskeleton-interacting proteins showed CML24 binding to the IQ domain of myosin VIII, but not to actin filaments (Abu–Abied et al. 2006). This is especially interesting in regard to chloroplast movement, as cell organelles are typically moved along actin filaments by myosin (Holweg et al. 2004; Wang et al. 2004). These disparate results suggest that CML24 has multiple roles in various tissues. Reports of a disturbed actin cytoskeleton in CML24 mutants could point to a role for CML24 in chloroplast anchoring, which is mediated by Ca2+/calmodulin and actin in spinach M cells (Takamatsu et al. 2011). As the PCPs are associated with actin, it is possible that CML24 acts as a chloroplast anchoring mediator in Bienertia. CML24 is also reported to effect microtubule organization, which is important for stabilizing the CC in Bienertia. A phosphatase 2C (PP2C) family protein was also identified (Table 3.9). PP2Cs are the largest phosphatase family in plants and involved in the regulation of signaling pathways. More specifically, they are suggested to act as negative regulators involved in ABA signaling (Schweighofer et al. 2004; Xue et al. 2008; Ma et al. 2009). The gene discovered here is the *B. vulgaris* homolog to Arabidopsis Clade E-Growth-Regulating PP2C-2 (EGR2) that acts as a negative regulator for plant growth in response to drought. It localizes to the cell periphery in Arabidopsis (Bhaskara et al. 2016). One of the identified targets of EGR2 is the Microtubule Associated Stress Protein 1 (MASP1) (Bhaskara et al. 2016). MASP1 is among the subset of genes that are overall more abundant in Bienertia than Suaeda. It is expressed at medium levels in Bienertia and undetected in Suaeda (Table A.1). It was shown to promote growth by stabilizing microtubules. The authors inferred that EGR2 and MASP1 work together to adjust plant growth to external signals (Bhaskara et al. 2016). MASP1 was also found alongside CML24 in a tube growth study (Wang et al. 2008)

and was identified as $C_4$ co-regulated in a $C_3/C_4$ transcriptomics study (Bräutigam et al. 2011). Microtubules might play a role in the positioning of BS cell chloroplasts in the Kranz $C_4$ species *E. coracana*, as disruption of the actin filaments has no effect on the positioning of BS chloroplasts (Kobayashi et al. 2009; Kandasamy et al. 1999). This makes the reported co-regulation of the microtubule stabilizing protein MASP1 with $C_4$ genes in mature *Cleome gynandra* leaves interesting, as it could point to a role in anchoring the stationary BS chloroplasts (Bräutigam et al. 2011). Taken together, the expression of EGR2 and MASP1 in Bienertia cotyledons suggests a role in the regulation and stability of microtubules, which could contribute to establishing $SCC_4$ morphology. The IQ-domain 2 protein (IQD2) was also identified in the screening of genes of interest (Table 3.9) and showed a 3.6-fold upregulation from TP1 to 3 in Bienertia. IQD2 belongs to a plant-specific family of calmodulin-binding proteins localizing to various subcellular domains, including microtubules, plasma membrane and nucleus. A role as scaffolding protein was inferred from accumulated findings (Bürstenbinder et al. 2017). Scaffolding proteins stabilize protein interactions, which makes them important contributors to many signaling pathways (Kundu et al. 2013; Teichert et al. 2014). IQD2 appears to localize to microtubules. Overexpression of IQD16 led to an altered cell shape and the cortical orientation of microtubules, while overexpression of two other IQD family proteins influenced plant growth. The authors propose a role for IQD family proteins in the regulation of cell shape, function and plant growth (Bürstenbinder et al. 2017). Taken together, IQD2 plays a role in organizing microtubule orientation. It is thus possible that it contributes to the formation of microtubule structures around the CC in Bienertia. The genes identified here, IQD2, CML24 and MASP1, are all reported to effect microtubule organisation, while the protein phosphatase EGR2 is identified as a regulator of MASP1. Taken together, this suggests that these genes are involved in the correct development of the microtubule network in Bienertia. While chloroplasts are not usually associated with microtubules, the CC in Bienertia cells is surrounded by a cage-like microtubule structure (Park et al. 2009). Therefore, the identified genes could potentially regulate the formation of the CC. In addition, CML24 could mediate anchoring of the PCPs via actin.

**Actin 7 is highly expressed in Bienertia cotyledons**

The cytoskeleton consists of multiple elements. While microtubules are implicated in establishing the orientation of cellular growth along one axis, organelles are dis-

tributed using the actin filament network (Lloyd et al. 1985; Bannigan et al. 2006; van Gestel et al. 2002; Nebenführ et al. 1999). Actin7 was also identified as possible effector of SCC$_4$ morphology (Table 3.9, Section 3.5.4). Arabidopsis ACT7 has been shown to be highest expressed in young plant tissues, including hypocotyl and seed coat. Its promotor contains phytohormone-responsive elements, which is a unique trait among actin isoforms (McDowell et al. 1996). Its expression is altered following auxin, cytokinin and ABA treatment and a role in cell specification of root epidermis is reported (McDowell et al. 1996; Kandasamy et al. 2001; Kandasamy et al. 2009). Expression data from etiolated Arabidopsis seedlings suggest a link between ACT7 and cell elongation (McDowell et al. 1996). In summary, a single actin isoform, ACT7, is more abundant in Bienertia than Suaeda cotyledons. Previous research on ACT7 in Arabidopsis revealed a role in general tissue development, cell differentiation and elongation, likely under regulation from phytohormones. The development of the actin network could be crucial for the chloroplast positioning process, given that organelles are typically moved along actin filaments for positioning and Bienertia PCPs were shown be associated with actin (Kandasamy et al. 1999; van Gestel et al. 2002). In addition, it was proposed that actin filaments support cytoplasmic strands (Hoffmann et al. 2004) , which are a documented feature of Bienertia chlorenchyma cells (Park et al. 2009). Therefore, ACT7 could be involved in cell differentiation or chloroplast positioning in Bienertia chlorenchyma cells.

Here, it was shown that several genes within both cytoskeletal networks (actin and microtubules) are differently expressed in dark-grown cotyledons of Bienertia and Suaeda. While previous research already indicated a potential role of the cytoskeleton in Bienertia chlorenchyma cell differentiation and the formation of the unique subcellular compartmentalization in SCC$_4$ species, this thesis provides for the first time direct candidates. Loss-of-function analysis can be used in the future to validate the functional relevance of the identified candidates.

### 4.1.2 Chloroplast movement genes might contribute to chloroplast anchoring in Bienertia

Chloroplast movement is a mechanism common to all plants. The process can be separated into light-dependent and light-independent (developmental) movement (Augustynowicz et al. 2003). The former gives plants the ability to react to various light-

conditions and optimize photosynthetic output and avoid damage by either maximizing the chloroplast surface in low light conditions or reducing it in high light conditions (Suetsugu et al. 2007; Kasahara et al. 2002). The latter is a less studied process which is responsible for distributing chloroplast to their "home position". Example for this are the even distribution in M cells of most $C_3$ plants, distribution to specific sides of the cell in the case of Kranz $C_4$ BS cells and *S. aralocaspica* chlorenchyma cells, or to the CC and PC in Bienertia. It is not clear whether the distributing mechanism is the same in these examples. Light-dependent movement is much better studied and several genes involved in light-dependent chloroplast movement have been identified. However, the phenotypes of loss-of-function mutants of some of these genes suggest that they might also play a role in light-independent chloroplast distribution and could thus mediate distribution to the CC and PC in Bienertia.

CHUP1 interacts with the chloroplast outer membrane and is necessary for chloroplast anchoring to the plasma membrane via a coiled-coil region in Arabidopsis (Oikawa et al. 2008). CHUP1 also contains an actin binding domain and a profilin binding region in the C-terminal part, which is suggested to mediate chloroplast movement (Oikawa et al. 2003; Schmidt von Braun et al. 2008). A loss-of-function mutant shows chloroplast aggregation at the bottom of cells (Oikawa et al. 2003). Expression of a truncated GFP-fusion protein containing the N-terminal and coiled-coil region, but lacking the actin and profilin binding domain leads to the loss of position for PCPs (Figure 3.21). This suggest that the coiled-coil region is not sufficient to mediate chloroplast anchoring to the plasma membrane in Bienertia. However, it could not be verified that the actin- and profilin binding domains are necessary for chloroplast positioning, as transient expression of a actin binding domain containing construct of CHUP1 was not successful. It is also possible that a different protein mediates chloroplast anchoring in Bienertia. Overexpression of the CHUP1-GFP protein could lead to displacement of the anchoring protein from the outer envelope membrane, as both compete for the same position on the outer envelope. Overexpression of outer envelope protein 7 (OEP7) in Arabidopsis caused chloroplast aggregation similar to the *chup1* knock-out mutant (Oikawa et al. 2008). Therefore, the data does not conclusively show that CHUP1 is involved in chloroplast anchoring in Bienertia. CHUP1 does not contain a transit peptide, instead, it was suggested that the N-terminus of CHUP1 inserts directly into the chloroplast outer membrane, possibly by recognizing a specific lipid composition (Oikawa et al. 2008). The truncated GFP-fusion protein, which contains the N-terminus, localizes to the chloroplast membrane. As the PCPs

aggregate around the CC in cells expressing the truncated protein, it could not be verified whether CHUP1 localizes specifically to the PCPs (Figure 3.21). In a proteomic study of the dimorphic chloroplasts, which identified numerous differentially accumulating proteins, CHUP1 was not quantified (Offermann et al. 2015). It thus remains unclear whether CHUP1 localizes exclusively to the PCPs. Taken together, it can not be excluded that CHUP1 is involved in anchoring the PCPs to the plasma membrane in Bienertia, but does not interact with the CCPs. Due to the proposed localizing mechanism, this would require the PCPs to have a different outer envelope lipid composition than CCPs. Here, the CHUP1 expression pattern in dark-grown Bienertia cotyledons is similar to that of many $C_4$ genes and expression increases after light exposure (Figure 3.22). Interestingly, CHUP1 is already highly expressed at TP1, before a separation of the chloroplast types is evident. This could indicate that individual plastids are retained at the plasma membrane by interacting with CHUP1, while the rest aggregates to form the pre-CC. It is unclear, how this interaction is mediated, as the plastids appear identical at this developmental stage.

Light-dependent chloroplast movement in Arabidopsis is further mediated by two kinesin-like proteins for actin-based chloroplast movement (KAC) (Suetsugu et al. 2010). KACs are shown to localize to the plasma membrane and the cytosol and interact with the chloroplast only indirectly via actin (Suetsugu et al. 2010; Vanstraelen et al. 2006). Double knockout mutants of KAC lack the chloroplast movement response in high- or low light conditions and additionally show detachment of chloroplasts from the plasma membrane (Suetsugu et al. 2010). KAC1 is expressed at low levels in Bienertia and Suaeda, while KAC2 is expressed at high levels (Figure 3.22). This is in contrast to expression in mature leaves in Arabidopsis, where KAC1 is expressed higher than KAC2 (Suetsugu et al. 2010). There is some indication that KAC2 is responsible for maintaining chloroplast positioning in the dark, as the mutant shows chloroplast accumulation at the cell bottom in darkness (Suetsugu et al. 2010). KAC2 has a two- to threefold higher overall expression in Suaeda compared to Bienertia (Figure 3.22). The higher expression in Suaeda could be explained by the presumably higher number of chloroplasts in contact with the plasma membrane, as most chloroplasts in Bienertia are located in the CC. However, this is in contrast with the expression of CHUP1, which is slightly higher in Bienertia. In summary, the high expression of KAC2 in contrast to KAC1 observed here might indicate that KAC2 plays a role in chloroplast positioning in developing dark-grown tissues. This is supported by the observation of aberrant chloroplast positioning under dark conditions

in the Arabidopsis KAC2 mutant. However, this has not been tested yet functionally. Like CHUP1, KAC2 is highly expressed at TP1, which might indicate that both genes collaborate in PCP anchoring.

Taken together, the expression profiles of CHUP1 and KAC2 and transient expression of a CHUP1-GFP fusion construct suggest that both genes are involved in chloroplast anchoring in Bienertia. Similar expression patterns in Suaeda indicate that the genes perform a role in chloroplast positioning in Suaeda as well. Interestingly, both CHUP1 and KAC2 are highly expressed at TP1, before the spatial separation of the plastids occur. This can be explained by the following scenario: Individual plastids are retained at the plasma membrane while the rest aggregates to form the pre-CC due to lack of interaction with CHUP1 and KAC2.

### 4.1.3 Transcriptional regulation of $SCC_4$ development

In addition to finding effectors of chloroplast positioning, this study also aimed to identify transcription factors involved in $SCC_4$ morphology development. Transcription factors play an important role in development and cell differentiation as they can modulate the expression of multiple genes. They often respond to hormonal or environmental stimuli. Phytohormone crosstalk is the main mechanism for regulating developmental processes (Arc et al. 2013; McAtee et al. 2013; Kohli et al. 2013). ABA is one of the main phytohormones and regulates many plant growth and development processes as well as responses to environmental factors (Cutler et al. 2010; Bray 2002; Léon–Kloosterziel et al. 1996; Assmann 2003).

**ABA signaling is potentially connected to $SCC_4$ development**

High ABA is generally thought to be connected to growth inhibition, however, in young tissues, high levels of ABA actually promote growth (Finkelstein 2013). In the $C_4$ plant finger millet, M cell chloroplast aggregative movement caused by stress is reportedly ABA mediated (Yamada et al. 2009; Maai et al. 2011). Aggregated M cell chloroplasts move towards the BS cell and form a cluster, which looks similar to the first steps of Bienertia CC development. The amphibious plant *Eleocharis vivipara* develops Kranz anatomy and performs $C_4$ photosynthesis in terrestrial conditions, but develops $C_3$ photosynthetic traits in submerged tissues (Ueno et al. 1988). ABA treat-

ment induced the formation of Kranz anatomy and $C_4$ gene expression in submerged tissues (Ueno 1998).

Several of the identified transcription factors are reported to be involved in ABA signaling. BEL-1-like homeodomain 1 (BLH1) is upregulated throughout dark-grown development in Bienertia, but not Suaeda (Table 3.10). Homeodomain proteins are essential for plant development by regulating cell differentiation and patterning (Baima et al. 2001; Haecker et al. 2004). The homeodomain-TALE-BEL family transcription factor BLH1 is important for the modulation of the ABA response in seedlings (Kim et al. 2013). Ectopic expression of BLH1 in the embryo sac leads among other effects to a disturbed nuclear localization (Pagnussat et al. 2007). The authors suggest that this might be caused by misregulation of the microtubule network. In Bienertia, BLH1 could potentially be involved in ABA-regulated microtubule organization. ABA-insensitive 4 (ABI4) is approximately sevenfold more abundant in dark-grown cotyledons of Bienertia compared to Suaeda (Table 3.10). It belongs to the APETALA2/ ethylene-responsive element binding proteins (AP2/EREBP) family of transcription factors. Apart from being involved in ABA signaling, there is some evidence that ABI4 also plays a role in chloroplast retrograde signaling (León et al. 2013). It was shown that ABI4 can bind to a regulatory unit of RuBisCO and act as a negative regulator of photosynthetic gene expression (Acevedo–Hernández et al. 2005). Other targets of ABI4 are the light-harvesting complex chlorophyll a/b protein of photosystem II (Lhcb) genes (Koussevitzky et al. 2007; Staneloni et al. 2008). This appears to be in contrast to the results shown here, which show high expression of RuBisCO and 11 of 14 identified Lhcb genes in Bienertia (Figure 3.13 and Table A.2). However, it was suggested that other transcription factors work redundantly with ABI4, as the loss-of-function mutant shows only transient effects on Lhcb gene expression (Koussevitzky et al. 2007). It seems that expression of ABI4 does not have an inhibitory effect on RuBisCO and Lhcb gene expression in dark-grown Bienertia cotyledons. The role of ABI4 and the integration of ABA signaling in $SCC_4$ development is unclear. Inducer of CBF expression 1 (ICE1) expression is fivefold increased in Bienertia compared to Suaeda (Table 3.10). ICE1 belongs to the bHLH transcription factor family and is reported to be a negative regulator of ABI4 and ABA-dependent responses in general (Liang et al. 2015). In addition, it was found to be involved in stomatal cell differentiation and cold sensing (Kanaoka et al. 2008; Chinnusamy et al. 2003). It was suggested that the different functions of ICE1 are controlled by post-translational modifications and protein interactions (Liang et al. 2015). Without knowledge of these

control mechanisms, it is not clear whether increased ICE1 expression in dark-grown Bienertia cotyledons negates the effects of ABI4 and ABA-dependent processes or whether it performs a different function. Overall, the limited knowledge on the individual transcription factors along with the complexity of ABA responses, which are manifold in all plant tissues and show a lot of cross-regulation with other signaling pathways, makes it difficult to infer a function for these genes from in silico information only. Functional studies in Bienertia are necessary for elucidating the role of ABA signaling in chlorenchyma cell development. In addition, a comparison of the data presented here to RNA-seq data from true leaf development could separate the transcription factors with a role in seedling development or skotomorphogenesis and pinpoint the ones regulating chlorenchyma development. Virtually all transcription factors and their families described here are involved in some aspect of plant development and the functional studies used to generate this knowledge are necessarily tissue- or process-specific, which makes identifying the ones involved in chloroplast positioning or chlorenchyma development a difficult task. This is not an uncommon occurrence in RNA-seq experiments. In maize, more than 1000 transcription factors potentially involved in Kranz $C_4$ tissue-specification have been identified, yet only four have so far been experimentally validated (Huang et al. 2016).

## 4.2 Dark-grown Bienertia cotyledons are a valid system for identification of SCC$_4$ factors

SCC$_4$ morphology develops in both dark- and light-grown cotyledons in Bienertia. Here, dark-grown cotyledons were chosen as the experimental setup for identifying factors involved in SCC$_4$ development. In light-grown plants, photosynthetic tissues undergo photomorphogenesis, however, at the same time, light-independent processes such as root development and cell differentiation take place. One of the advantages of using dark-grown tissue is that the developmentally regulated processes are uncoupled from the light-regulated ones, which reduces complexity of the various developmental processes. The most highly abundant genes in green tissue are photosynthetic genes, which are often light-regulated (Ma, 2001, Schmid, 2005). However, the genes responsible for SCC$_4$ morphology development in the dark are unlikely to be either photosynthetic or light-regulated. When designing this experiment, it was hypothesized that by restricting the light-regulated pathways and thus preven-

ting expression of the usually highly abundant photosynthetic genes, the sequencing depth for potential genes of interest could be increased. In etiolated cotyledons of both Bienertia and Suaeda, photosynthetic genes constitute 4-9 % of the highest expressed genes. In *C. gynandra*, it was shown that the percentage of photosynthetic genes in light-grown seedlings is higher and constitutes more than 30 % (Külahoglu et al. 2014). This shows that the sequencing depth for non-photosynthetic genes has indeed been increased by using dark-grown cotyledons. The downside to studying dark-grown development is that there is a very limited number of other studies on skotomorphogenesis, while cotyledon development in light-grown seedlings is a widely studied process. Nevertheless, dark-grown cotyledons can provide useful information on light-independent development by reducing complexity of signaling crosstalk and increasing sequencing depth for potential $SCC_4$ factors.

Cotyledons are the first leaves of a seedling and are functionally similar to true leaves, despite having a different origin. They develop during embryogenesis and not from the stem apical meristem. Here, cotyledons were chosen to investigate the developmentally controlled aspects of the $SCC_4$ phenomenon. It is not always the case that cotyledons of $C_4$ species perform $C_4$ photosynthesis, e.g. in the species *Salsola gemmascens*, *S. soda* (Pyankov et al. 2000; Lauterbach et al. 2017) and *H. ammodendron* (Li et al. 2015), cotyledons show a $C_3$ signature. However, Bienertia cotyledons contain the typical $SCC_4$ chlorenchyma cells and have been shown to perform $C_4$ photosynthesis (Freitag et al. 2002; Akhani et al. 2005). $SCC_4$ morphology also develops in dark-grown cotyledons (Figure 3.2). A previous study examined light-independent leaf development on enclosed branches of a light-grown plant and found that the two subcellular domains in newly formed leaves are developed in a similar way to light-grown leaves (Lara et al. 2008). However, long-distance signaling between the light- and dark-grown branches can not be excluded. Experiments have shown that exposure to light not only regulates gene expression in the light-exposed tissues, but can also induce gene expression in roots (Hemm et al. 2004; Correll et al. 2005; Usami et al. 2004; Dyachok et al. 2011; Shin et al. 2010; Bai et al. 2014). Cotyledons are therefore better suited than leaves to study the light-independent development of $SCC_4$.

It can thus be concluded that cotyledons are suitable to explore the aspects of development that they share with the primary photosynthetic tissue, especially $SCC_4$ morphology, and are especially useful for experiments relying on dark conditions. The main disadvantage is that cotyledon development in light has been investigated

much more frequently, and thus could have provided better resources for comparison to other species.

## 4.3 SCC$_4$ development in etiolated Bienertia cotyledons comprises more than chloroplast positioning

### SCC$_4$-specific chloroplast positioning occurs in etiolated Bienertia cotyledons

Chloroplast positioning in the SCC$_4$ species Bienertia and S. aralocaspica is essential for establishment of the C$_4$ pathway by creating separate locations for the primary and secondary CO$_2$ fixation (Voznesenskaya et al. 2001; Jurić et al. 2017). In Kranz C$_4$ species, unusual chloroplast positioning is observed in BS cells, where the centrifugal or centripetal location, depending on C$_4$ sub-type, is suggested to reduce CO$_2$ leakage or enhance metabolite flow between M and BS cells (Hattersley et al. 1981). Chlorenchyma cells in dark-grown Bienertia cotyledons develop the typical SCC$_4$ morphology in a light-independent way by spatially separating the plastids into a ball-like CC and single plastids distributed around the cell periphery (Figure 3.2). In Suaeda aralocaspica, another SCC$_4$ species, separation of chloroplast types also occurs in dark-grown cotyledons (Voznesenskaya et al. 2004). In some Kranz C$_4$ species, e.g. *E. coracana*, *A. hypochondriacus*, and maize, BS cell chloroplasts are positioned correctly in etiolated plants (Miyake et al. 1987; Wang et al. 1993; Taniguchi et al. 2003). In other Kranz C$_4$ species, e.g. in *E. utilis*, *S. bicolor* and *E. aristidea*, the positioning is disturbed in dark-grown seedlings (Taniguchi et al. 2003). Accordingly, there is a light-independent chloroplast positioning mechanism in both SCC$_4$ and some Kranz C$_4$ species, although the details and its regulation are still unresolved. The divergent behaviour in other Kranz C$_4$ species suggests that it is not a conserved process and C$_4$ plants have found different ways to regulate chloroplast positioning.

### Several C$_4$-related genes are highly expressed in dark-grown cotyledons of Bienertia compared to Suaeda

C$_4$ evolved from C$_3$ photosynthesis by recruiting genes into the C$_4$ pathway that fulfill a different role in C$_3$ species (Aubry et al. 2011). As a result, the genes acquired new regulatory features and show a general increase in expression (Hibberd et al. 2010;

Ku et al. 1996). Like many photosynthetic genes, most $C_4$ genes show increased expression in response to light (Burgess et al. 2016; Sheen et al. 1987). This thesis shows that most $C_4$ genes are more abundant in Bienertia compared to Suaeda and are even among the most abundant genes of dark-grown Bienertia cotyledons (Figure 3.13). In dark-grown cotyledons of *S. aralocaspica*, the $C_4$ genes PPDK and PEPC were not detected, while RuBisCO expression increased over time (Voznesenskaya et al. 2004). In etiolated maize seedlings, high expression of eight $C_4$ genes were found, including PEPC (Xu et al. 2016). This is in contrast to earlier studies, which found low expression of PPDK and PEPC (Sheen et al. 1987) as well as NADP-ME and NADP-MDH (Langdale et al. 1988) in etiolated maize seedlings. In *C. gynandra*, most $C_4$ genes are expressed in dark-grown tissues, but show upregulation in light (Burgess et al. 2016). In *Flaveria trinervia*, low levels of PPDK and PEPC transcripts were detected as well as expression of Rubisco SU (Shu et al. 1999). The expression levels of $C_4$ genes in the dark differ across species, but not $C_4$ types. Bienertia (SCC$_4$) , maize (NADP-ME) and *C. gynandra* (NAD-ME) show high levels of $C_4$ gene expression, while *S. aralocaspica* (SCC$_4$) and *F. trinervia* (NAP-ME) express $C_4$ genes at low levels in the dark. This implies that $C_4$ gene expression in the former species is at least partly under developmental control. The regulatory mechanism behind this has not yet been investigated.

**Are Bienertia etioplasts capable of differential protein accumulation?**

A crucial step in $C_4$ development is the differential protein accumulation in the correct cell- (in Kranz $C_4$ species) or chloroplast type (in SCC$_4$ species). PPDK and PEPC localize to M cells in Kranz $C_4$ species and the PC in Bienertia, while RuBisCO and NAD-ME are exclusively found in BS cells of Kranz $C_4$ species and the CC in Bienertia (Langdale et al. 1991; Voznesenskaya et al. 2002). In *A. hypochondriacus*, the development of $C_4$ is largely light-independent (Wang et al. 1993). In dark-grown cotyledons, PEPC, PPDK and RuBisCO are localized to the appropriate cell types, although expression levels are lower than in light-grown seedlings. However, the data does not show whether PPDK and RuBisCO are imported correctly into the etioplasts (Wang et al. 1993). In Bienertia, several proteins that were shown to accumulate differentially in dimorphic chloroplasts (e. g. PPDK, BASS, RuBisCO) (Offermann et al. 2015), are highly expressed in etiolated Bienertia cotyledons (Figure 3.13). This raises the interesting question whether Bienertia etioplasts are capable of differentially accumulation of these proteins. The mechanism by which differential accumulation works is

still under investigation, however, a recent study has shown that motifs within the chloroplast transit peptides are responsible for import into the PCPs. The authors hypothesized that components connected to the import machinery might recognize the motifs and block import into the CCCP (Wimmer et al. 2017). While etioplasts differ from chloroplasts in regards to protein composition and ultrastructure and undergo massive changes during de-etiolation (Kleffmann et al. 2007; López-Juez 2007), the composition of the import machinery is not necessarily different. Rice etioplasts contain high levels of complete import complexes, presumably to allow for increased import needs during de-etiolation (Reiland et al. 2011). It is thus possible that the prerequisites for selective accumulation are already present in etioplasts. Whether or not Bienertia etioplasts are capable of differential protein accumulation is an intriguing question that could provide valuable insights into the mechanism underlying differential protein accumulation and should be investigated further. Dark-grown seedlings provide a suitable experimental setup in which this question can be addressed.

Etiolated Bienertia cotyledons show light-independent $C_4$ development that encompasses not just differential plastid positioning, but also high expression of several $C_4$ genes and potentially differential protein accumulation. Both differential plastid positioning and elevated $C_4$ gene expression were also found in etiolated maize and *A. hypochondriacus* seedlings, but were not always observed in other Kranz $C_4$ species. The high variety of $C_4$ development in the dark among Kranz $C_4$ plants is perhaps not surprising, considering that $C_4$ evolved independently at least 61 times (Sage, 2016). This suggests that for some $C_4$ species, part of the $C_4$ regulation is developmentally and light-independently controlled, giving even etiolated tissues of $C_4$ plants a "$C_4$-footprint". This phenomenon has been largely neglected by researchers focussing on the greening process for identifying regulators for selective expression of genes in M and BS cells that might not be light-controlled. Studies on etiolated tissues could identify light-independent regulators of $C_4$ photosynthesis by separating the factors controlling developmental gene expression from the factors involved in light-regulated gene expression.

## 4.4 Evaluation of the technical aspects of comparative transcriptomics

**Cross-species mapping is a useful tool in comparative transcriptomics studies**

Cross-species mapping has been successfully used in several studies concerned with finding $C_4$-related genes by sequencing closely-related $C_3$ and $C_4$ species and $C_3$-$C_4$ intermediates (Bräutigam et al. 2011; Gowik et al. 2011; Külahoglu et al. 2014). These studies are based on the assumption that most genes behave in a similar way in both species and only $C_4$-related genes are differentially expressed. Here, the $SCC_4$ species Bienertia was compared to the $C_3$ species Suaeda with the aim to identify factors in $SCC_4$ development and chloroplast positioning in particular. Several cross-species mapping strategies were tested here in order to make sure that the expression data of both species can be reliably compared to each other (Figure 3.3). Ultimately, the employed mapping strategy was to use a reference species that is closely related to both sequenced plants (*B. vulgaris*) and map the reads in protein space, i.e. translated reads in all six reading frames (Section 3.3.1). Here, it was shown that there is no mapping bias that favors one species over the other. The CLC mapping approach with strict parameters for both species was skewed towards Bienertia, while loosening the mapping parameters for Suaeda led to a shift towards higher expression values in Suaeda (Figure 3.3). This pattern can likely be explained by mismatches in read mapping caused by relaxed mapping parameters. A positive side-effect of using *B. vulgaris* as a reference species was access to a well-curated transcriptome of a sequenced species (Dohm et al. 2014). Measurement of selected gene expression by qPCR matched well with the mapping data (Figure 3.17). However, mappings to very similar isoforms can be incorrectly quantified, as shown for actin (Section 3.5.4). This should be taken into consideration when evaluating the data. In summary, it can be concluded that the cross-species mapping approach was successful and the workflow established here can be used for further comparative transcriptomics studies.

**Evaluation of methods used for identifying $SCC_4$ factors**

In order to find the genes responsible for $SCC_4$ morphology, several methods were employed. *k*-means clustering was used to find genes with increasing expression in Bienertia and little change in Suaeda (Figure 3.16). This approach was also used to

identify $C_4$-related expression changes, regulatory genes and evolutionary changes in a comparative study of $C_3$, $C_4$, and $C_3$-$C_4$ intermediate species of the genus *Flaveria* (Gowik et al. 2011). In addition, strict filtering criteria were chosen based on the assumption that genes involved in this process are upregulated in sync with the observed changes in plastid positioning. Given that altered gene expression can change cell morphology and cell differentiation, the assumption is reasonable (Camilleri et al. 2002; Köllmer et al. 2014). In addition, combining observed morphological differences with gene expression data in Kranz and non-Kranz maize leaves led to the discovery of potential regulators of Kranz anatomy (Wang et al. 2013). Moreover, the average gene expression across three developmental stages served as a basis to find genes more abundant in one species over the other (Table A.1). Potential transcriptional regulators of $SCC_4$ development were compiled from a combination of the methods described here (Table 3.10). The genes identified here comprise a comprehensive resource of potential $SCC_4$ development effectors and regulators that can serve as a basis for functional analysis.

This is the first study to use a transcriptomics approach in the $SCC_4$ species Bienertia to elucidate factors involved in the development of $SCC_4$ morphology. To allow for a comparison to the $C_3$ species Suaeda, several read mapping approaches were tested to determine the best for a cross-species comparison. In the end, a workflow for protein-space based mapping to the reference transcriptome of *B. vulgaris* was successfully established. Candidate genes were identified using a combination of various bioinformatical methods.

## 4.5 Conclusion

This thesis provides a detailed analysis on cotyledon development in the $SCC_4$ species Bienertia and identifies genes involved in $SCC_4$ morphology. These can serve as resource for functional analysis and future research. The role of the cytoskeleton in $SCC_4$ development was solidified and a role for ABA signaling is proposed. In addition, this is the first comparative transcriptomics study of skotomorphogenesis in cotyledons of a $C_4$ and $C_3$ species. A workflow for comparative transcriptomics for Bienertia was established that can be replicated and adjusted for follow-up experiments. One of the advantages of RNA sequencing is that the data can be reanalyzed in a different context and thus serve to answer more questions (Rung et al. 2013; Sun

et al. 2014). The relative high expression of some genes related to $C_4$ metabolism in dark grown cotyledons observed in this study suggests that more aspects of $C_4$ photosynthesis than just chloroplast positioning are regulated light-independently. Similar observations were made in some Kranz $C_4$ species which provides an interesting foundation for experiments on light-independent $C_4$ development.

# References

Abu–Abied, M., Golomb, L., Belausov, E., Huang, S., Geiger, B., Kam, Z., Staiger, C. J., and Sadot, E. (2006). Identification of plant cytoskeleton–interacting proteins by screening for actin stress fiber association in mammalian fibroblasts. *The Plant Journal* **48**: 367–379.

Acevedo–Hernández, G. J., León, P., and Herrera–Estrella, L. R. (2005). Sugar and ABA responsiveness of a minimal RBCS light–responsive unit is mediated by direct binding of ABI4. *The Plant Journal* **43**: 506–519.

Aida, M., Ishida, T., Fukaki, H., Fujisawa, H., and Tasaka, M. (1997). Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. *The Plant Cell* **9**: 841–857.

Akhani, H., Barroca, J., Koteeva, N., Voznesenskaya, E. V., Franceschi, V., Edwards, G., Ghaffari, S., and Ziegler, H. (2005). Bienertia sinuspersici (Chenopodiaceae): A New Species from Southwest Asia and Discovery of a Third Terrestrial C4 Plant Without Kranz Anatomy. *Systematic Botany* **30**: 290–301.

Akhani, H., Chatrenoor, T., Dehghani, M., Khoshravesh, R., Mahdavi, P., and Matin-zadeh, Z. (2012). A new species of Bienertia (Chenopodiaceae) from Iranian salt deserts: A third species of the genus and discovery of a fourth terrestrial C4 plant without Kranz anatomy. *Plant Biosystems* **146**: 550–559.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology* **11**: R106.

Arc, E., Sechet, J., Corbineau, F., Rajjou, L., and Marion-Poll, A. (2013). ABA crosstalk with ethylene and nitric oxide in seed dormancy and germination. *Frontiers in plant science* **4**: 63.

Armstrong, D. P. and Westoby, M. (1993). Seedlings from Large Seeds Tolerated Defoliation Better: A Test Using Phylogeneticaly Independent Contrasts. *Ecology* **74**: 1092–1100.

Assmann, S. M. (2003). OPEN STOMATA1 opens the door to ABA signaling in Arabidopsis guard cells. *Trends in Plant Science* **8**: 151–153.

Aubry, S., Brown, N. J., and Hibberd, J. M. (2011). The role of proteins in C3 plants prior to their recruitment into the C4 pathway. *Journal of Experimental Botany* **62**: 3049–3059.

Augustynowicz, J., Krzeszowiec, W., and Gabrys, H. (2003). Acquisition of plastid movement responsiveness to light during mesophyll cell differentiation. *International Journal of Developmental Biology* **53**: 121–127.

Babraham Bioinformatics (2017). FASTQC - A quality control tool for high throughput sequence data. `http://www.bioinformatics.babraham.ac.uk/projects/fastqc/`.

Bai, S., Yao, T., Li, M., Guo, X., Zhang, Y., Zhu, S., and He, Y. (2014). PIF3 is involved in the primary root growth inhibition of Arabidopsis induced by nitric oxide in the light. *Molecular plant* **7**: 616–625.

Baima, S., Possenti, M., Matteucci, A., Wisman, E., Altamura, M. M., Ruberti, I., and Morelli, G. (2001). The Arabidopsis ATHB-8 HD-Zip Protein Acts as a Differentiation-Promoting Transcription Factor of the Vascular Meristems. *Plant Physiology* **126**: 643–655.

Bannigan, A., Wiedemeier, A. M. D., Williamson, R. E., Overall, R. L., and Baskin, T. I. (2006). Cortical Microtubule Arrays Lose Uniform Alignment Between Cells and are Oryzalin Resistant in the Arabidopsis Mutant, radially swollen 6. *Plant and Cell Physiology* **47**: 949–958.

Barrero, R. A., Chapman, B., Yang, Y., Moolhuijzen, P., Keeble-Gagnère, G., Zhang, N., Tang, Q., Bellgard, M. I., and Qiu, D. (2011). De novo assembly of Euphorbia fischeriana root transcriptome identifies prostratin pathway related genes. *BMC Genomics* **12**: 600.

Beemster, G. T., Veylder, L. D., Vercruysse, S., West, G., Rombaut, D., van Hummelen, P., Galichet, A., Gruissem, W., Inzé, D., and Vuylsteke, M. (2005). Genome-Wide Analysis of Gene Expression Profiles Associated with Cell Cycle Transitions in Growing Organs of Arabidopsis. *Plant Physiology* **138**: 734–743.

Besseau, S., Hoffmann, L., Geoffroy, P., Lapierre, C., Pollet, B., and Legrand, M. (2007). Flavonoid Accumulation in Arabidopsis Repressed in Lignin Synthesis Affects Auxin Transport and Plant Growth. *The Plant Cell* **19**: 148–162.

Bhaskara, G. B., Wen, T.-N., Nguyen, T. T., and Verslues, P. E. (2016). Protein Phosphatase 2Cs and Microtubule-Associated Stress Protein 1 Control Microtubule Stability, Plant Growth, and Drought Response. *The Plant Cell* **29**: 169–191.

Blilou, I., Frugier, F., Folmer, S., Serralbo, O., Willemsen, V., Wolkenfelt, H., Eloy, N. B., Ferreira, P. C., Weisbeek, P., and Scheres, B. (2002). The Arabidopsis HOBBIT gene encodes a CDC27 homolog that links the plant cell cycle to progression of cell differentiation. *Genes & Development* **16**: 2566–2575.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Boyd, C. N., Franceschi, V. R., Chuong, S. D., Akhani, H., Kiirats, O., Smith, M., and Edwards, G. E. (2007). Flowers of Bienertia cycloptera and Suaeda aralocaspica (Chenopodiaceae) complete the life cycle performing single-cell C 4 photosynthesis. *Functional Plant Biology* **34**: 268.

Bräutigam, A., Kajala, K., Wullenweber, J., Sommer, M., Gagneul, D., Weber, K. L., Carr, K. M., Gowik, U., Maß, J., Lercher, M. J., Westhoff, P., Hibberd, J. M., and Weber, A. P. (2011). An mRNA Blueprint for C4 Photosynthesis Derived from Comparative Transcriptomics of Closely Related C3 and C4 Species. *Plant Physiology* **155**: 142–156.

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

Bray, E. A. (2002). Abscisic acid regulation of gene expression during water–deficit stress in the era of the Arabidopsis genome. *Plant, cell & environment* **25**: 153–161.

Brown, N. J., Newell, C. A., Stanley, S., Chen, J. E., Perrin, A. J., Kajala, K., and Hibberd, J. M. (2011). Independent and Parallel Recruitment of Preexisting Mechanisms Underlying C4 Photosynthesis. *Science* **331**: 1436–1439.

Burgess, S. J., Granero-Moya, I., Grangé-Guermente, M. J., Boursnell, C., Terry, M. J., and Hibberd, J. M. (2016). Ancestral light and chloroplast regulation form the foundations for C4 gene expression. *Nature Plants* **2**: 16161.

Bürstenbinder, K., Möller, B., Plötner, R., Stamm, G., Hause, G., Mitra, D., and Abel, S. (2017). The IQD Family of Calmodulin-Binding Proteins Links Calcium Signaling to Microtubules, Membrane Microdomains, and the Nucleus. *Plant Physiology* **173**: 1692–1708.

Camilleri, C., Azimzadeh, J., Pastuglia, M., Bellini, C., Grandjean, O., and Bouchez, D. (2002). The Arabidopsis TONNEAU2 Gene Encodes a Putative Novel Protein Phosphatase 2A Regulatory Subunit Essential for the Control of the Cortical Cytoskeleton. *The Plant Cell* **14**: 833–845.

Carolin, R. C., Jacobs, S. W., and Vesk, M. (1978). Kranz cells and mesophyll in the Chenopodiales. *Australian Journal of Botany* **26**: 683.

Chen, M., Galvão, R. M., Li, M., Burger, B., Bugea, J., Bolado, J., and Chory, J. (2010). Arabidopsis HEMERA/pTAC12 Initiates Photomorphogenesis by Phytochromes. *Cell* **141**: 1230–1240.

Chen, Y., Yin, H., Gao, M., Zhu, H., Zhang, Q., and Wang, Y. (2016). Comparative Transcriptomics Atlases Reveals Different Gene Expression Pattern Related to Fusarium Wilt Disease Resistance and Susceptibility in Two Vernicia Species. *Frontiers in plant science* **7**: 1974.

Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B.-h., Hong, X., Agarwal, M., and Zhu, J.-K. (2003). ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. *Genes & Development* **17**: 1043–1054.

Chuong, S. D., Franceschi, V. R., and Edwards, G. E. (2006). The Cytoskeleton Maintains Organelle Partitioning Required for Single-Cell C4 Photosynthesis in Chenopodiaceae Species. *The Plant Cell* **18**: 2207–2223.

Ciftci-Yilmaz, S. and Mittler, R. (2008). The zinc finger network of plants. *Cellular and Molecular Life Sciences* **65**: 1150–1160.

Colbert, J. T. (1991). Regulation of type I phytochrome mRNA abundance. *Physiologia Plantarum* **82**: 327–332.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., and Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology* **17**: 13.

Correll, M. J. and Kiss, J. Z. (2005). The Roles of Phytochromes in Elongation and Gravitropism of Roots. *Plant and Cell Physiology* **46**: 317–323.

Covshoff, S., Szecowka, M., Hughes, T. E., Smith-Unna, R., Kelly, S., Bailey, K. J., Sage, T. L., Pachebat, J. A., Leegood, R., and Hibberd, J. M. (2016). C4 Photosynthesis in the Rice Paddy: Insights from the Noxious Weed Echinochloa glabrescens. *Plant Physiology* **170**: 57–73.

Cutler, S. R., Rodriguez, P. L., Finkelstein, R. R., and Abrams, S. R. (2010). Abscisic acid: emergence of a core signaling network. *Annual review of plant biology* **61**: 651–679.

Dai, X., Sinharoy, S., Udvardi, M., and Zhao, P. X. (2013). PlantTFcat: an online plant transcription factor and transcriptional regulator categorization and analysis tool. *BMC Bioinformatics* **14**: 321.

Delk, N. A., Johnson, K. A., Chowdhury, N. I., and Braam, J. (2005). CML24, Regulated in Expression by Diverse Stimuli, Encodes a Potential Ca2+ Sensor That Functions

in Responses to Abscisic Acid, Daylength, and Ion Stress. *Plant Physiology* **139**: 240–253.

Dengler, N. G., Dengler, R. E., Donnelly, P. M., and Hattersley, P. W. (1994). Quantitative Leaf Anatomy of C3 and C4 Grasses (Poaceae): Bundle Sheath and Mesophyll Surface Area Relationships. *Annals of Botany* **73**: 241–255.

Ding, Z., Weissmann, S., Wang, M., Du, B., Huang, L., Wang, L., Tu, X., Zhong, S., Myers, C., Brutnell, T. P., Sun, Q., and Li, P. (2015). Identification of Photosynthesis-Associated C4 Candidate Genes through Comparative Leaf Gradient Transcriptome in Multiple Lineages of C3 and C4 Species. *PLOS ONE* **10**: e0140629.

Dohm, J. C., Minoche, A. E., Holtgräwe, D., Capella-Gutiérrez, S., Zakrzewski, F., Tafer, H., Rupp, O., Sörensen, T. R., Stracke, R., Reinhardt, R., Goesmann, A., Kraft, T., Schulz, B., Stadler, P. F., Schmidt, T., Gabaldón, T., Lehrach, H., Weisshaar, B., and Himmelbauer, H. (2014). The genome of the recently domesticated crop plant sugar beet (Beta vulgaris). *Nature* **505**: 546–549.

Dyachok, J., Zhu, L., Liao, F., He, J., Huq, E., and Blancaflor, E. B. (2011). SCAR Mediates Light-Induced Root Elongation in Arabidopsis through Photoreceptors and Proteasomes. *The Plant Cell* **23**: 3610–3626.

Earley, K. W., Haag, J. R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C. S. (2006). Gateway–compatible vectors for plant functional genomics and proteomics. *The Plant Journal* **45**: 616–629.

Edwards, D., Murray, J. A., and Smith, A. G. (1998). Multiple Genes Encoding the Conserved CCAAT-Box Transcription Factor Complex Are Expressed in Arabidopsis. *Plant Physiology* **117**: 1015–1022.

Ehleringer, J. R., Sage, R. F., Flanagan, L. B., and Pearcy, R. W. (1991). Climate change and the evolution of C4 photosynthesis. *Trends in Ecology & Evolution* **6**: 95–99.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences* **95**: 14863–14868.

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**: 953–971.

Finkelstein, R. (2013). Abscisic Acid Synthesis and Response. *The Arabidopsis Book / American Society of Plant Biologists* **11**: e0166.

Freitag, H. and Stichler, W. (2000). A Remarkable New Leaf Type With Unusual Photosynthetic Tissue in a Central Asiatic Genus of Chenopodiaceae. *Plant Biology* **2**: 154–160.

Freitag, H. and Stichler, W. (2002). Bienertia cycloptera Bunge ex Boiss., Chenopodiaceae, another C4 Plant without Kranz Tissues. *Plant Biology* **4**: 121–132.

Furumoto, T., Yamaguchi, T., Ohshima-Ichie, Y., Nakamura, M., Tsuchida-Iwata, Y., Shimamura, M., Ohnishi, J., Hata, S., Gowik, U., Westhoff, P., Bräutigam, A., Weber, A. P., and Izui, K. (2011). A plastidial sodium-dependent pyruvate transporter. *Nature* **476**: 472–475.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In: *The Proteomics Protocols Handbook*. Ed. by J. M. Walker. Totowa, NJ: Humana Press Inc, pp. 571–607. ISBN: 978-1-59259-890-8. DOI: 10.1385/1-59259-890-0:571.

Gibbons, F. D. and Roth, F. P. (2002). Judging the Quality of Gene Expression-Based Clustering Methods Using Gene Annotation. *Genome Research* **12**: 1574–1581.

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.

Haecker, A., Gross-Hardt, R., Geiges, B., Sarkar, A., Breuninger, H., Herrmann, M., and Laux, T. (2004). Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in Arabidopsis thaliana. *Development* **131**: 657–668.

Hanson, J., Johannesson, H., and Engström, P. (2001). Sugar-dependent alterations in cotyledon and leaf development in transgenic plants expressing the HDZhdip gene ATHB13. *Plant Molecular Biology* **45**: 247–262.

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.

Hattersley, P. W. and Browning, A. J. (1981). Occurrence of the suberized lamella in leaves of grasses of different photosynthetic types. I. In parenchymatous bundle sheaths and PCR ("Kranz") sheaths. *Protoplasma* **109**: 371–401.

Hemm, M. R., Rider, S. D., Ogas, J., Murry, D. J., and Chapple, C. (2004). Light induces phenylpropanoid metabolism in Arabidopsis roots. *The Plant Journal* **38**: 765–778.

Hibberd, J. M. and Covshoff, S. (2010). The regulation of gene expression required for C4 photosynthesis. *Annual review of plant biology* **61**: 181–207.

Himmelbauer, H. and Weisshaar, B. (2017). The Beta vulgaris Resource: Sugar beet genetics and genomics. `http://bvseq.molgen.mpg.de`.

Hoffmann, A. and Nebenführ, A. (2004). Dynamic rearrangements of transvacuolar strands in BY-2 cells imply a role of myosin in remodeling the plant actin cytoskeleton. *Protoplasma* **224**: 201–210.

Holweg, C. and Nick, P. (2004). Arabidopsis myosin XI mutant is defective in organelle movement and polar auxin transport. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 10488–10493.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008). Genevestigator V3: A Reference Expression Database for the Meta-Analysis of Transcriptomes. *Advances in Bioinformatics* **2008**: 420747.

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**: 1–13.

Huang, P. and Brutnell, T. P. (2016). A synthesis of transcriptomic surveys to dissect the genetic basis of C4 photosynthesis. *Current Opinion in Plant Biology* **31**: 91–99.

Hyun, Y. and Lee, I. (2006). KIDARI, encoding a non-DNA Binding bHLH protein, represses light signal transduction in Arabidopsis thaliana. *Plant Molecular Biology* **61**: 283–296.

Ikeda, M., Fujiwara, S., Mitsuda, N., and Ohme-Takagi, M. (2012). A Triantagonistic Basic Helix-Loop-Helix System Regulates Cell Elongation in Arabidopsis. *The Plant Cell* **24**: 4483–4497.

Jiao, Y., Ma, L., Strickland, E., and Deng, X. W. (2005). Conservation and Divergence of Light-Regulated Genome Expression Patterns during Seedling Development in Rice and Arabidopsis. *The Plant Cell* **17**: 3239–3256.

Jurić, I., González-Pérez, V., Hibberd, J. M., Edwards, G., and Burroughs, N. J. (2017). Size matters for single-cell C4 photosynthesis in Bienertia. *Journal of Experimental Botany* **68**: 255–267.

Kadota, A., Yamada, N., Suetsugu, N., Hirose, M., Saito, C., Shoda, K., Ichikawa, S., Kagawa, T., Nakano, A., and Wada, M. (2009). Short actin-based mechanism for light-directed chloroplast movement in Arabidopsis. *Proceedings of the National Academy of Sciences* **106**: 13106–13111.

Kagawa, T., Sakai, T., Suetsugu, N., Oikawa, K., Ishiguro, S., Kato, T., Tabata, S., Okada, K., and Wada, M. (2001). Arabidopsis NPL1: A Phototropin Homolog Controlling the Chloroplast High-Light Avoidance Response. *Science* **291**: 2138–2141.

Kanaoka, M. M., Pillitteri, L. J., Fujii, H., Yoshida, Y., Bogenschutz, N. L., Takabayashi, J., Zhu, J.-K., and 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.

Kandasamy, M. K., Gilliland, L. U., McKinney, E. C., and Meagher, R. B. (2001). One Plant Actin Isovariant, ACT7, Is Induced by Auxin and Required for Normal Callus Formation. *The Plant cell* **13**: 1541–1554.

Kandasamy, M. K., McKinney, E. C., and Meagher, R. B. (2009). A Single Vegetative Actin Isovariant Overexpressed under the Control of Multiple Regulatory Sequences Is Sufficient for Normal Arabidopsis Development. *The Plant cell* **21**: 701–718.

Kandasamy, M. K. and Meagher, R. B. (1999). Actin–organelle interaction: Association with chloroplast in Arabidopsis leaf mesophyll cells. *Cytoskeleton* **44**: 110–118.

Kanervo, E., Singh, M., Suorsa, M., Paakkarinen, V., Aro, E., Battchikova, N., and Aro, E.-M. (2008). Expression of protein complexes and individual proteins upon transition of etioplasts to chloroplasts in pea (Pisum sativum). *Plant & cell physiology* **49**: 396–410.

Kasahara, M., Kagawa, T., Oikawa, K., Suetsugu, N., Miyao, M., and Wada, M. (2002). Chloroplast avoidance movement reduces photodamage in plants. *Nature* **420**: 829–832.

Kent, W. J. (2002). BLAT–the BLAST-like alignment tool. *Genome research* **12**: 656–664.

Kim, D., Cho, Y., Ryu, H., Kim, Y., Kim, T., and Hwang, I. (2013). BLH1 and KNAT3 modulate ABA responses during germination and early seedling development in Arabidopsis. *The Plant Journal* **75**: 755–766.

King, J. L., Edwards, G. E., and Cousins, A. B. (2012). The efficiency of the CO2–concentrating mechanism during single–cell C4 photosynthesis. *Plant, cell & environment* **35**: 513–523.

Kirst, M., Caldo, R., Casati, P., Tanimoto, G., Walbot, V., Wise, R. P., and Buckler, E. S. (2006). Genetic diversity contribution to errors in short oligonucleotide microarray analysis. *Plant Biotechnology Journal* **4**: 489–498.

Kleffmann, T., von Zychlinski, A., Russenberger, D., Hirsch-Hoffmann, M., Gehrig, P., Gruissem, W., and Baginsky, S. (2007). Proteome Dynamics during Plastid Differentiation in Rice. *Plant Physiology* **143**: 912–923.

Kobayashi, H., Yamada, M., Taniguchi, M., Kawasaki, M., Sugiyama, T., and Miyake, H. (2009). Differential Positioning of C4 Mesophyll and Bundle Sheath Chloroplasts: Recovery of Chloroplast Positioning Requires the Actomyosin System. *Plant and Cell Physiology* **50**: 129–140.

Kohli, A., Sreenivasulu, N., Lakshmanan, P., and Kumar, P. P. (2013). The phytohormone crosstalk paradigm takes center stage in understanding how plants respond to abiotic stresses. *Plant Cell Reports* **32**: 945–957.

Köllmer, I., Novák, O., Strnad, M., Schmülling, T., and Werner, T. (2014). Overexpression of the cytosolic cytokinin oxidase/dehydrogenase (CKX7) from Arabidopsis causes specific changes in root growth and xylem differentiation. *The Plant Journal* **78**: 359–371.

Koncz, C. and Schell, J. (1986). The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. *Molecular and General Genetics MGG* **204**: 383–396.

Kong, S.-G., Kagawa, T., Wada, M., and Nagatani, A. (2013). A C-Terminal Membrane Association Domain of Phototropin 2 is Necessary for Chloroplast Movement. *Plant and Cell Physiology* **54**: 57–68.

Königer, M. and Bollinger, N. (2012). Chloroplast movement behavior varies widely among species and does not correlate with high light stress tolerance. *Planta* **236**: 411–426.

Königer, M., Jessen, B., Yang, R., Sittler, D., and Harris, G. C. (2010). Light, genotype, and abscisic acid affect chloroplast positioning in guard cells of Arabidopsis thaliana leaves in distinct ways. *Photosynthesis Research* **105**: 213–227.

Koteyeva, N. K., Voznesenskaya, E. V., Berry, J. O., Cousins, A. B., and Edwards, G. E. (2016). The unique structural and biochemical development of single cell C4 photosynthesis along longitudinal leaf gradients in Bienertia sinuspersici and Suaeda aralocaspica (Chenopodiaceae). *Journal of Experimental Botany* **67**: 2587–2601.

Koussevitzky, S., Nott, A., Mockler, T. C., Hong, F., Sachetto-Martins, G., Surpin, M., Lim, J., Mittler, R., and Chory, J. (2007). Signals from Chloroplasts Converge to Regulate Nuclear Gene Expression. *Science* **316**: 715–719.

Ku, M. S., Kano-Murakami, Y., and Matsuoka, M. (1996). Evolution and expression of C4 photosynthesis genes. *Plant physiology* **111**: 949–957.

Külahoglu, C., Denton, A. K., Sommer, M., Maß, J., Schliesky, S., Wrobel, T. J., Berckmans, B., Gongora-Castillo, E., Buell, C. R., Simon, R., Veylder, L. D., Bräutigam, A., and Weber, A. P. (2014). Comparative Transcriptome Atlases Reveal Altered Gene Expression Modules between Two Cleomaceae C3 and C4 Plant Species. *The Plant Cell* **26**: 3243–3260.

Kundu, N., Dozier, U., Deslandes, L., Somssich, I. E., and Ullah, H. (2013). Arabidopsis scaffold protein RACK1A interacts with diverse environmental stress and photosynthesis related proteins. *Plant Signaling & Behavior* **8**: e24012.

Kwong, R. W., Bui, A. Q., Lee, H., Kwong, L. W., Fischer, R. L., Goldberg, R. B., and Harada, J. J. (2003). LEAFY COTYLEDON1-LIKE Defines a Class of Regulators Essential for Embryo Development. *The Plant Cell* **15**: 5–18.

Langdale, J. A., Zelitch, I., Miller, E., and Nelson, T. (1988). Cell position and light influence C4 versus C3 patterns of photosynthetic gene expression in maize. *The EMBO Journal* **7**: 3643–3651.

Langdale, J. A. and Nelson, T. (1991). Spatial regulation of photosynthetic development in C4 plants. *Trends in Genetics* **7**: 191–196.

Lara, M. V., Offermann, S., Smith, M., Okita, T. W., Andreo, C. S., and Edwards, G. E. (2008). Leaf development in the single-cell C4 system in Bienertia sinuspersici: expression of genes and peptide levels for C4 metabolism in relation to chlorenchyma structure under different light conditions. *Plant physiology* **148**: 593–610.

Lauterbach, M., Billakurthi, K., Kadereit, G., Ludwig, M., Westhoff, P., and Gowik, U. (2017). C3 cotyledons are followed by C4 leaves: intra-individual transcriptome analysis of Salsola soda (Chenopodiaceae). *Journal of Experimental Botany* **68**: 161–176.

Leivar, P., Monte, E., Oka, Y., Liu, T., Carle, C., Castillon, A., Huq, E., and Quail, P. H. (2008). Multiple Phytochrome-Interacting bHLH Transcription Factors Repress Premature Seedling Photomorphogenesis in Darkness. *Current Biology* **18**: 1815–1823.

León, P., Gregorio, J., and Cordoba, E. (2013). ABI4 and its role in chloroplast retrograde communication. *Frontiers in Plant Science* **3**: 304.

Léon–Kloosterziel, K. M., Gil, M. A., Ruijs, G. J., Jacobsen, S. E., Olszewski, N. E., Schwartz, S. H., Zeevaart, J. A., and Koornneef, M. (1996). Isolation and characterization of abscisic acid–deficient Arabidopsis mutants at two new loci. *The Plant Journal* **10**: 655–661.

Li, Y., Ma, X., Zhao, J., Xu, J., Shi, J., Zhu, X.-G., Zhao, Y., and Zhang, H. (2015). Developmental Genetic Mechanisms of C4 Syndrome Based on Transcriptome Analysis of C3 Cotyledons and C4 Assimilating Shoots in Haloxylon ammodendron. *PLOS ONE* **10**: e0117175.

Liang, C.-H. and Yang, C.-C. (2015). Identification of ICE1 as a negative regulator of ABA-dependent pathways in seeds and seedlings of Arabidopsis. *Plant Molecular Biology* **88**: 459–470.

Listiawan, D. A., Tanoue, R., Kobayashi, K., and Masuda, T. (2015). Expression Analysis of Transcription Factors Involved in Chloroplast Differentiation. *Procedia Chemistry* **14**: 146–151.

Liu, Q., Wang, Z., Xu, X., Zhang, H., and Li, C. (2015). Genome-Wide Analysis of C2H2 Zinc-Finger Family Transcription Factors and Their Responses to Abiotic Stresses in Poplar (Populus trichocarpa). *PLOS ONE* **10**: e0134753.

Lloyd, C. W., Clayton, L., Dawson, P. J., Doonan, J. H., Hulme, J. S., Roberts, I. N., and Wells, B. (1985). The Cytoskeleton Underlying Side Walls and Cross Walls in Plants: Molecules and Macromolecular Assemblies. *Journal of Cell Science* **1985**: 143–155.

Lohse, M., Bolger, A. M., Nagel, A., Fernie, A. R., Lunn, J. E., Stitt, M., and Usadel, B. (2012). RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research* **40**: W622–W627.

Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., Tohge, T., Fernie, A. R., Stitt, M., and Usadel, B. (2014). Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, cell & environment* **37**: 1250–1258.

López-Juez, E. (2007). Plastid biogenesis, between light and shadows. *Journal of Experimental Botany* **58**: 11–26.

Lotan, T., Ohto, M.-a., Yee, K. M., West, M. A., Lo, R., Kwong, R. W., Yamagishi, K., Fischer, R. L., Goldberg, R. B., and Harada, J. J. (1998). Arabidopsis LEAFY COTYLEDON1 Is Sufficient to Induce Embryo Development in Vegetative Cells. *Cell* **93**: 1195–1205.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550.

Lu, Y., Huggins, P., and Bar-Joseph, Z. (2009). Cross species analysis of microarray expression data. *Bioinformatics* **25**: 1476–1483.

Lulin, H., Xiao, Y., Pei, S., Wen, T., and Shangqin, H. (2012). The First Illumina-Based De Novo Transcriptome Sequencing and Analysis of Safflower Flowers. *PLOS ONE* **7**: e38653.

Lupas, A., van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. *Science (New York, N.Y.)* **252**: 1162–1164.

Ma, Y., Szostkiewicz, I., Korte, A., Moes, D., Yang, Y., Christmann, A., and Grill, E. (2009). Regulators of PP2C Phosphatase Activity Function as Abscisic Acid Sensors. *Science* **324**: 1064–1068.

Maai, E., Shimada, S., Yamada, M., Sugiyama, T., Miyake, H., and Taniguchi, M. (2011). The avoidance and aggregative movements of mesophyll chloroplasts in C4 monocots in response to blue light and abscisic acid. *Journal of Experimental Botany* **62**: 3213–3221.

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

Maloof, J. N., Borevitz, J. O., Dabi, T., Lutes, J., Nehring, R. B., Redfern, J. L., Trainer, G. T., Wilson, J. M., Asami, T., Berry, C. C., Weigel, D., and Chory, J. (2001). Natural variation in light sensitivity of Arabidopsis. *Nature Genetics* **29**: 441–446.

Mathews, S. (2006). Phytochrome–mediated development in land plants: red light sensing evolves to meet the challenges of changing light environments. *Molecular Ecology* **15**: 3483–3503.

McAtee, P., Karim, S., Schaffer, R., and David, K. (2013). A dynamic interplay between phytohormones is required for fruit development, maturation, and ripening. *Frontiers in plant science* **4**: 79.

McDowell, J. M., An, Y., Huang, S., McKinney, E. C., and Meagher, R. B. (1996). The arabidopsis ACT7 actin gene is expressed in rapidly developing tissues and responds to several external stimuli. *Plant physiology* **111**: 699–711.

Miyake, H. and Yamamoto, Y. (1987). Centripetal Disposition of Bundle Sheath Chloroplasts During the Leaf Development of Eleusine coracana. *Annals of Botany* **60**: 641–647.

Miyoshi, K., Ito, Y., Serizawa, A., and Kurata, N. (2003). OsHAP3 genes regulate chloroplast biogenesis in rice. *The Plant Journal* **36**: 532–540.

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.

Mol, J., Grotewold, E., and Koes, R. (1998). How genes paint flowers and seeds. *Trends in Plant Science* **3**: 212–217.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* **5**: 621–628.

Muhaidat, R., Sage, R. F., and Dengler, N. G. (2007). Diversity of Kranz anatomy and biochemistry in C4 eudicots. *American Journal of Botany* **94**: 362–381.

Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A., Vonrhein, C., Moras, D., Romier, C., Bolognesi, M., and Mantovani, R. (2013). Sequence-Specific Transcription Factor NF-Y Displays Histone-like DNA Binding and H2B-like Ubiquitination. *Cell* **152**: 132–143.

Nebenführ, A., Gallagher, L. A., Dunahay, T. G., Frohlick, J. A., Mazurkiewicz, A. M., Meehl, J. B., and Staehelin, L. A. (1999). Stop-and-Go Movements of Plant Golgi Stacks Are Mediated by the Acto-Myosin System. *Plant Physiology* **121**: 1127–1141.

Offermann, S., Friso, G., Doroshenk, K. A., Sun, Q., Sharpe, R. M., Okita, T. W., Wimmer, D., Edwards, G. E., and van Wijk, K. J. (2015). Developmental and Subcellular Organization of Single-Cell C$_4$ Photosynthesis in Bienertia sinuspersici Determined by Large-Scale Proteomics and cDNA Assembly from 454 DNA Sequencing. *Journal of proteome research* **14**: 2090–2108.

Offermann, S., Okita, T. W., and Edwards, G. E. (2011). Resolving the compartmentation and function of C4 photosynthesis in the single-cell C4 species Bienertia sinuspersici. *Plant physiology* **155**: 1612–1628.

Ogle, K. (2003). Implications of interveinal distance for quantum yield in C4 grasses: a modeling and meta-analysis. *Oecologia* **136**: 532–542.

Ohashi-Ito, K., Oguchi, M., Kojima, M., Sakakibara, H., and Fukuda, H. (2013). Auxin-associated initiation of vascular cell differentiation by LONESOME HIGHWAY. *Development* **140**: 765–769.

Oikawa, K., Kasahara, M., Kiyosue, T., Kagawa, T., Suetsugu, N., Takahashi, F., Kanegae, T., Niwa, Y., Kadota, A., and Wada, M. (2003). CHLOROPLAST UNUSUAL POSITIONING1 Is Essential for Proper Chloroplast Positioning. *The Plant Cell* **15**: 2805–2815.

Oikawa, K., Yamasato, A., Kong, S.-G., Kasahara, M., Nakai, M., Takahashi, F., Ogura, Y., Kagawa, T., and Wada, M. (2008). Chloroplast Outer Envelope Protein CHUP1 Is Essential for Chloroplast Anchorage to the Plasma Membrane and Chloroplast Movement. *Plant Physiology* **148**: 829–842.

Oshlack, A., Chabot, A. E., Smyth, G. K., and Gilad, Y. (2007). Using DNA microarrays to study gene expression in closely related species. *Bioinformatics* **23**: 1235–1242.

Osterlund, M. T., Hardtke, C. S., Wei, N., and Deng, X. W. (2000). Targeted destabilization of HY5 during light-regulated development of Arabidopsis. *Nature* **405**: 462–466.

Pagnussat, G. C., Yu, H.-J., and Sundaresan, V. (2007). Cell-Fate Switch of Synergid to Egg Cell in Arabidopsis eostre Mutant Embryo Sacs Arises from Misexpression of the BEL1-Like Homeodomain Gene BLH1. *The Plant Cell* **19**: 3578–3592.

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2016). The impact of amplification on differential expression analyses by RNA-seq. *Scientific reports* **6**: 25533.

Park, J., Edwards, G. E., Knoblauch, M., and Okita, T. W. (2009). Structural changes in the vacuole and cytoskeleton are key to development of the two cytoplasmic domains supporting single-cell C4 photosynthesis in Bienertia sinuspersici. *Planta* **229**: 369–382.

Prigge, M. J. and Wagner, D. R. (2001). The Arabidopsis SERRATE Gene Encodes a Zinc-Finger Protein Required for Normal Shoot Development. *The Plant Cell* **13**: 1263–1280.

Pyankov, V. I., Voznesenskaya, E. V., Kuzmin, A. N., Ku, M. S. B., Ganko, E., Franceschi, V. R., Black, C. C., and Edwards, G. E. (2000). Occurrence of C3 and C4 photosynthesis in cotyledons and leaves of Salsola species (Chenopodiaceae). *Photosynthesis Research* **63**: 69–84.

Qiu, J.-L., Jilk, R., Marks, M. D., and Szymanski, D. B. (2002). The Arabidopsis SPIKE1 Gene Is Required for Normal Cell Shape Control and Tissue Development. *The Plant Cell* **14**: 101–118.

R Core Team (2017). R: A Language and Environment for Statistical Computing. `https://www.R-project.org/`.

Rao, X., Lu, N., Li, G., Nakashima, J., Tang, Y., and Dixon, R. A. (2016). Comparative cell-specific transcriptomics reveals differentiation of C4 photosynthesis pathways in switchgrass and other C4 lineages. *Journal of Experimental Botany* **67**: 1649–1662.

Reed, J. W., Nagatani, A., Elich, T. D., Fagan, M., and Chory, J. (1994). Phytochrome A and Phytochrome B Have Overlapping but Distinct Functions in Arabidopsis Development. *Plant Physiology* **104**: 1139–1149.

Reiland, S., Grossmann, J., Baerenfaller, K., Gehrig, P., Nunes–Nesi, A., Fernie, A. R., Gruissem, W., and Baginsky, S. (2011). Integrated proteome and metabolite analysis of the de–etiolation process in plastids from rice (Oryza sativa L.) *Proteomics* **11**: 1751–1763.

Rung, J. and Brazma, A. (2013). Reuse of public genome-wide gene expression data. *Nature Reviews Genetics* **14**: 89–99.

Runge, S., Sperling, U., Frick, G., Apel, K., and Armstrong, G. A. (1996). Distinct roles for light–dependent NADPH:protochlorophyllide oxidoreductases (POR) A and B during greening in higher plants. *The Plant Journal* **9**: 513–523.

Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., and Quackenbush, J. (2003). TM4: a free, open-source system for microarray data management and analysis. *BioTechniques* **34**: 374–378.

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

Sage, R. F. (2016). A portrait of the C 4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *Journal of Experimental Botany* **67**: 4039–4056.

Sakai, T., Kagawa, T., Kasahara, M., Swartz, T. E., Christie, J. M., Briggs, W. R., Wada, M., and Okada, K. (2001). Arabidopsis nph1 and npl1: Blue light receptors that mediate both phototropism and chloroplast relocation. *Proceedings of the National Academy of Sciences* **98**: 6969–6974.

Schmidt von Braun, S. and Schleiff, E. (2008). The chloroplast outer membrane protein CHUP1 interacts with actin and profilin. *Planta* **227**: 1151–1159.

Schütze, P., Freitag, H., and Weising, K. (2003). An integrated molecular and morphological study of the subfamily Suaedoideae Ulbr. (Chenopodiaceae). *Plant Systematics and Evolution* **239**: 257–286.

Schweighofer, A., Hirt, H., and Meskiene, I. (2004). Plant PP2C phosphatases: emerging functions in stress signaling. *Trends in Plant Science* **9**: 236–243.

Seluzicki, A., Burko, Y., and Chory, J. (2017). Dancing in the dark: darkness as a signal in plants. *Plant, cell & environment*. Advance online publication. doi: 10.1111/pce.12900.

Sheen, J. Y. and Bogorad, L. (1987). Differential expression of C4 pathway genes in mesophyll and bundle sheath cells of greening maize leaves. *Journal of Biological Chemistry* **262**: 11726–11730.

Shin, D. H., Cho, M.-H., Kim, T.-L., Yoo, J., Kim, J.-I., Han, Y.-J., Song, P.-S., Jeon, J.-S., Bhoo, S. H., and Hahn, T.-R. (2010). A small GTPase activator protein interacts with cytoplasmic phytochromes in regulating root development. *The Journal of biological chemistry* **285**: 32151–32159.

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 C4 Plant, Flaveria trinervia. *Plant Physiology* **121**: 731–741.

Sigrist, C. J. A., Castro, E. de, Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Research* **41**: D344–D347.

Šlajcherová, K., Fišerová, J., Fischer, L., and Schwarzerová, K. (2012). Multiple Actin Isotypes in Plants: Diverse Genes for Diverse Roles? *Frontiers in plant science* **3**: 226.

Slewinski, T. L., Anderson, A. A., Zhang, C., and Turgeon, R. (2012). Scarecrow Plays a Role in Establishing Kranz Anatomy in Maize Leaves. *Plant and Cell Physiology* **53**: 2030–2037.

Solymosi, K. and Böddi, B. (2006). Optical properties of bud scales and protochlorophyll(ide) forms in leaf primordia of closed and opened buds. *Tree Physiology* **26**: 1075–1085.

Solymosi, K., Martinez, K., Kristóf, Z., Sundqvist, C., and Böddi, B. (2004). Plastid differentiation and chlorophyll biosynthesis in different leaf layers of white cabbage (Brassica oleracea cv. capitata). *Physiologia Plantarum* **121**: 520–529.

Staneloni, R. J., Rodriguez-Batiller, M. J., and Casal, J. J. (2008). Abscisic Acid, High-Light, and Oxidative Stress Down-Regulate a Photosynthetic Gene via a Promoter Motif Not Involved in Phytochrome-Mediated Transcriptional Regulation. *Molecular plant* **1**: 75–83.

Strand, Å., Asami, T., Alonso, J., Ecker, J. R., and Chory, J. (2003). Chloroplast to nucleus communication triggered by accumulation of Mg-protoporphyrinIX. *Nature* **421**: 79–83.

Suetsugu, N., Sato, Y., Tsuboi, H., Kasahara, M., Imaizumi, T., Kagawa, T., Hiwatashi, Y., Hasebe, M., and Wada, M. (2012). The KAC Family of Kinesin-Like Proteins is Essential for the Association of Chloroplasts with the Plasma Membrane in Land Plants. *Plant and Cell Physiology* **53**: 1854–1865.

Suetsugu, N. and Wada, M. (2007). Chloroplast photorelocation movement mediated by phototropin family proteins in green plants. *Biological Chemistry* **388**: 927–935.

Suetsugu, N., Yamada, N., Kagawa, T., Yonekura, H., Uyeda, T. Q. P., Kadota, A., and Wada, M. (2010). Two kinesin-like proteins mediate actin-based chloroplast movement in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* **107**: 8860–8865.

Sun, X., Yang, Q., Deng, Z., and Ye, X. (2014). Digital Inventory of Arabidopsis Transcripts Revealed by 61 RNA Sequencing Samples. *Plant Physiology* **166**: 869–878.

Sun, X., Zhou, S., Meng, F., and Liu, S. (2012). De novo assembly and characterization of the garlic (Allium sativum) bud transcriptome by Illumina sequencing. *Plant Cell Reports* **31**: 1823–1828.

Surpin, M., Larkin, R. M., and Chory, J. (2002). Signal transduction between the chloroplast and the nucleus. *The Plant cell* **14**: S327–38.

Takamatsu, H. and Takagi, S. (2011). Actin-Dependent Chloroplast Anchoring is Regulated by Ca2+–Calmodulin in Spinach Mesophyll Cells. *Plant and Cell Physiology* **52**: 1973–1982.

Tanaka, A. and Tsuji, H. (1985). Appearance of Chlorophyll-Protein Complexes in Greening Barley Seedlings. *Plant and Cell Physiology* **26**: 893–902.

Taniguchi, Y., Taniguchi, M., Kawasaki, M., and Miyake, H. (2003). Strictness of the Centrifugal Location of Bundle Sheath Chloroplasts in Different NADP-ME Type C 4 Grasses. *Plant Production Science* **6**: 274–280.

Teichert, I., Steffens, E. K., Schnaß, N., Fränzel, B., Krisp, C., Wolters, D. A., and Kück, U. (2014). PRO40 Is a Scaffold Protein of the Cell Wall Integrity Pathway, Linking the MAP Kinase Module to the Upstream Activator Protein Kinase C. *PLOS Genetics* **10**: e1004582.

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L. A., Rhee, S. Y., and 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.

Thirumurugan, T., Ito, Y., Kubo, T., Serizawa, A., and Kurata, N. (2008). Identification, characterization and interaction of HAP family genes in rice. *Molecular Genetics and Genomics* **279**: 279–289.

Tholen, D., Boom, C., Noguchi, K., Ueda, S., Katase, T., and Terashima, I. (2008). The chloroplast avoidance response decreases internal conductance to CO2 diffusion in Arabidopsis thaliana leaves. *Plant, cell & environment* **31**: 1688–1700.

Tian, Z.-D., Zhang, Y., Liu, J., and Xie, C.-H. (2010). Novel potato C2H2–type zinc finger protein gene, StZFP1, which responds to biotic and abiotic stress, plays a role in salt tolerance. *Plant Biology* **12**: 689–697.

Trojan, A. and Gabrys, H. (1996). Chloroplast Distribution in Arabidopsis thaliana (L.) Depends on Light Conditions during Growth. *Plant Physiology* **111**: 419–425.

Ueno, O. (1998). Induction of kranz anatomy and C4-like biochemical characteristics in a submerged amphibious plant by abscisic acid. *The Plant cell* **10**: 571–584.

Ueno, O., Samejima, M., Muto, S., and Miyachi, S. (1988). Photosynthetic characteristics of an amphibious plant, Eleocharis vivipara: Expression of C4 and C3 modes in contrasting environments. *Proceedings of the National Academy of Sciences* **85**: 6733–6737.

Usadel, B., Nagel, A., Steinhauser, D., Gibon, Y., Bläsing, O. E., Redestig, H., Sreenivasulu, N., Krall, L., Hannah, M. A., Poree, F., Fernie, A. R., and Stitt, M. (2006). PageMan: An interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. *BMC Bioinformatics* **7**: 535.

Usami, T., Mochizuki, N., Kondo, M., Nishimura, M., and Nagatani, A. (2004). Cryptochromes and Phytochromes Synergistically Regulate Arabidopsis Root Greening under Blue Light. *Plant and Cell Physiology* **45**: 1798–1808.

van Gestel, K., Köhler, R. H., and Verbelen, J.-P. (2002). Plant mitochondria move on F–actin, but their positioning in the cortical cytoplasm depends on both F–actin and microtubules. *Journal of Experimental Botany* **53**: 659–667.

van Noort, V., Snel, B., and Huynen, M. A. (2003). Predicting gene function by conserved co-expression. *Trends in Genetics* **19**: 238–242.

Vandepoele, K., Raes, J., Veylder, L. D., Rouzé, P., Rombauts, S., and Inzé, D. (2002). Genome-Wide Analysis of Core Cell Cycle Genes in Arabidopsis. *The Plant Cell* **14**: 903–916.

Vanstraelen, M., van Damme, D., de Rycke, R., Mylle, E., Inzé, D., and Geelen, D. (2006). Cell Cycle-Dependent Targeting of a Kinesin at the Plasma Membrane Demarcates the Division Site in Plant Cells. *Current Biology* **16**: 308–314.

Vlieghe, K., Boudolf, V., Beemster, G. T., Maes, S., Magyar, Z., Atanassova, A., Almeida Engler, J. de, Groodt, R. de, Inzé, D., and Veylder, L. D. (2005). The DP-E2F-like Gene DEL1 Controls the Endocycle in Arabidopsis thaliana. *Current Biology* **15**: 59–63.

von Arnim, A. and Deng, X.-W. (1996). Light control of seedling development. *Annual review of plant physiology and plant molecular biology* **47**: 215–243.

von Zychlinski, A., Kleffmann, T., Krishnamurthy, N., Sjölander, K., Baginsky, S., and Gruissem, W. (2005). Proteome Analysis of the Rice Etioplast Metabolic and Regulatory Networks and Novel Protein Functions. *Molecular & Cellular Proteomics* **4**: 1072–1084.

Voznesenskaya, E. V., Franceschi, V. R., and Edwards, G. E. (2004). Light-dependent development of single cell C4 photosynthesis in cotyledons of Borszczowia aralocaspica (Chenopodiaceae) during transformation from a storage to a photosynthetic organ. *Annals of botany* **93**: 177–187.

Voznesenskaya, E. V., Franceschi, V. R., Kiirats, O., Artyusheva, E. G., Freitag, H., and Edwards, G. E. (2002). Proof of C4 photosynthesis without Kranz anatomy in Bienertia cycloptera (Chenopodiaceae). *The Plant Journal* **31**: 649–662.

Voznesenskaya, E. V., Franceschi, V. R., Kiirats, O., Freitag, H., and Edwards, G. E. (2001). Kranz anatomy is not essential for terrestrial C4 plant photosynthesis. *Nature* **414**: 543–546.

Voznesenskaya, E. V., Koteyeva, N. K., Chuong, S. D., Akhani, H., Edwards, G. E., and Franceschi, V. R. (2005). Differentiation of cellular and biochemical features of the single-cell C4 syndrome during leaf development in Bienertia cycloptera (Chenopodiaceae). *American Journal of Botany* **92**: 1784–1795.

Wang, J. L., Long, J. J., Hotchkiss, T., and Berry, J. O. (1993). C4 Photosynthetic Gene Expression in Light- and Dark-Grown Amaranth Cotyledons. *Plant Physiology* **102**: 1085–1093.

Wang, L., Czedik-Eysenberg, A., Mertz, R. A., Si, Y., Tohge, T., Nunes-Nesi, A., Arrivault, S., Dedow, L. K., Bryant, D. W., Zhou, W., Xu, J., Weissmann, S., Studer, A., Li, P., Zhang, C., LaRue, T., Shao, Y., Ding, Z., Sun, Q., Patel, R. V., Turgeon, R., Zhu, X., Provart, N. J., Mockler, T. C., Fernie, A. R., Stitt, M., Liu, P., and Brutnell, T. P. (2014). Comparative analyses of C4 and C3 photosynthesis in developing leaves of maize and rice. *Nature Biotechnology* **32**: 1158–1165.

Wang, P., Kelly, S., Fouracre, J. P., and Langdale, J. A. (2013). Genome–wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C4 Kranz anatomy. *The Plant Journal* **75**: 656–670.

Wang, Y., Zhang, W.-Z., Song, L.-F., Zou, J.-J., Su, Z., and Wu, W.-H. (2008). Transcriptome Analyses Show Changes in Gene Expression to Accompany Pollen Germination and Tube Growth in Arabidopsis. *Plant Physiology* **148**: 1201–1211.

Wang, Y., Wang, B., Gilroy, S., Chehab, E. W., and Braam, J. (2011). CML24 is Involved in Root Mechanoresponses and Cortical Microtubule Orientation in Arabidopsis. *Journal of Plant Growth Regulation* **30**: 467–479.

Wang, Z. and Pesacreta, T. C. (2004). A subclass of myosin XI is associated with mitochondria, plastids, and the molecular chaperone subunit TCP-1alpha in maize. *Cell motility and the cytoskeleton* **57**: 218–232.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**: 57–63.

Wellburn, A. R. and Hampp, R. (1979). Appearance of photochemical function in prothylakoids during plastid development. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **547**: 380–397.

West, M., Yee, K. M., Danao, J., Zimmerman, J. L., Fischer, R. L., Goldberg, R. B., and Harada, J. J. (1994). LEAFY COTYLEDON1 Is an Essential Regulator of Late Embryogenesis and Cotyledon Identity in Arabidopsis. *The Plant Cell* **6**: 1731–1745.

Whatley, J. M. (1974). Chloroplast Development In Primary Leaves Of Phaseolus Vulgaris. *New Phytologist* **73**: 1097–1110.

Wimmer, D., Bohnhorst, P., Shekhar, V., Hwang, I., and Offermann, S. (2017). Transit peptide elements mediate selective protein targeting to two different types of chloroplasts in the single-cell C4 species Bienertia sinuspersici. *Scientific reports* **7**: 41187.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., and Provart, N. J. (2007). An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets. *PLOS ONE* **2**: e718.

Wu, L., Zhou, Z.-Y., Zhang, C.-G., Chai, J., Zhou, Q., Wang, L., Hirnerová, E., Mrvková, M., Novák, O., and Guo, G.-Q. (2015). Functional Roles of Three Cutin Biosynthetic Acyltransferases in Cytokinin Responses and Skotomorphogenesis. *PLOS ONE* **10**: e0121943.

Wurtzel, O., Sesto, N., Mellin, J. R., Karunker, I., Edelheit, S., Bécavin, C., Archambaud, C., Cossart, P., and Sorek, R. (2012). Comparative transcriptomics of pathogenic and non–pathogenic Listeria species. *Molecular Systems Biology* **8**: 583.

Xu, J., Bräutigam, A., Weber, A. P., and Zhu, X.-G. (2016). Systems analysis of cis-regulatory motifs in C 4 photosynthesis genes using maize and rice leaf transcriptomic data during a process of de-etiolation. *Journal of Experimental Botany* **67**: 5105–5117.

Xue, T., Wang, D., Zhang, S., Ehlting, J., Ni, F., Jakab, S., Zheng, C., and Zhong, Y. (2008). Genome-wide and expression analysis of protein phosphatase 2C in rice and Arabidopsis. *BMC Genomics* **9**: 550.

Yamada, M., Kawasaki, M., Sugiyama, T., Miyake, H., and Taniguchi, M. (2009). Differential Positioning of C4 Mesophyll and Bundle Sheath Chloroplasts: Aggregative Movement of C4 Mesophyll Chloroplasts in Response to Environmental Stresses. *Plant and Cell Physiology* **50**: 1736–1749.

Yan, A., Wu, M., Zhao, Y., Zhang, A., Liu, B., Schiefelbein, J., and Gan, Y. (2014). Involvement of C2H2 zinc finger proteins in the regulation of epidermal cell fate determination in Arabidopsis. *Journal of Integrative Plant Biology* **56**: 1112–1117.

Yang, X., Wang, S.-S., Wang, M., Qiao, Z., Bao, C.-C., and Zhang, W. (2014). Arabidopsis thaliana calmodulin-like protein CML24 regulates pollen tube growth by modulating the actin cytoskeleton and controlling the cytosolic Ca2+ concentration. *Plant Molecular Biology* **86**: 225–236.

Yi, C. and Deng, X. W. (2005). COP1 – from plant photomorphogenesis to mammalian tumorigenesis. *Trends in Cell Biology* **15**: 618–625.

Yu, C.-P., Chen, S. C.-C., Chang, Y.-M., Liu, W.-Y., Lin, H.-H., Lin, J.-J., Chen, H. J., Lu, Y.-J., Wu, Y.-H., Lu, M.-Y. J., Lu, C.-H., Shih, A. C.-C., Ku, M. S.-B., Shiu, S.-H., Wu, S.-H., and Li, W.-H. (2015). Transcriptome dynamics of developing maize leaves and genomewide prediction of cis elements and their cognate transcription factors. *Proceedings of the National Academy of Sciences* **112**: E2477–E2486.

Yun, J.-Y., Weigel, D., and Lee, I. (2002). Ectopic Expression of SUPERMAN Suppresses Development of Petals and Stamens. *Plant and Cell Physiology* **43**: 52–57.

Zenoni, S., Ferrarini, A., Giacomelli, E., Xumerle, L., Fasoli, M., Malerba, G., Bellin, D., Pezzotti, M., and Delledonne, M. (2010). Characterization of Transcriptional Complexity during Berry Development in Vitis vinifera Using RNA-Seq. *Plant Physiology* **152**: 1787–1795.

Zhang, T. and Folta, K. M. (2012). Green light signaling and adaptive response. *Plant Signaling & Behavior* **7**: 75–78.

Zhang, L.-Y., Bai, M.-Y., Wu, J., Zhu, J.-Y., Wang, H., Zhang, Z., Wang, W., Sun, Y., Zhao, J., Sun, X., Yang, H., Xu, Y., Kim, S.-H., Fujioka, S., Lin, W.-H., Chong, K., Lu, T., and Wang, Z.-Y. (2009). Antagonistic HLH/bHLH transcription factors mediate brassinosteroid regulation of cell elongation and plant development in rice and Arabidopsis. *The Plant cell* **21**: 3767–3780.

Zhiponova, M. K., Morohashi, K., Vanhoutte, I., Machemer-Noonan, K., Revalska, M., van Montagu, M., Grotewold, E., and Russinova, E. (2014). Helix–loop–helix/basic helix–loop–helix transcription factor network represses cell elongation in Arabidopsis through an apparent incoherent feed-forward loop. *Proceedings of the National Academy of Sciences* **111**: 2824–2829.

Zhu, S., Tang, S., Tan, Z., Yu, Y., Dai, Q., and Liu, T. (2017). Comparative transcriptomics provide insight into the morphogenesis and evolution of fistular leaves in Allium. *BMC Genomics* **18**: 60.

Zurzycki, J. (1955). Chloroplasts arrangement as a factor in photosynthesis. *Acta Societatis Botanicorum Poloniae* **24**: 27–63.

# A  Appendix

## List of electronic supplemental material

**Table 1**   Results of all read mapping scenarios

**Table 2**   List of all differentially expressed genes in Bienertia and Suaeda

**Table 3**   Results from $k$-means clustering of differentially expressed genes

**Table 4**   Additional information on candidate genes

**Table 5**   List of all genes with a fivefold higher expression in either species

**Table 6**   List of transcription factors with an expression above 20 NC in Bienertia or Suaeda

# A.1 Enrichment analysis
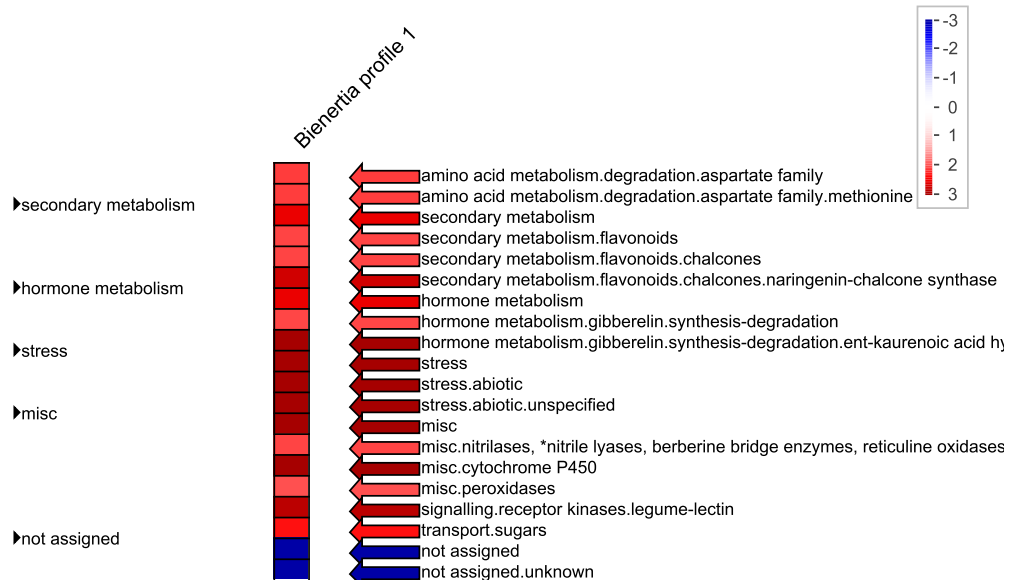


Figure A.1: Enrichment analysis of gene profiles with PageMan. The gene set was compared to all genes above the cutoff value using the PageMan module in MapMan. Red boxes show overrepresented BINs, blue boxes show underrepresented BINs.
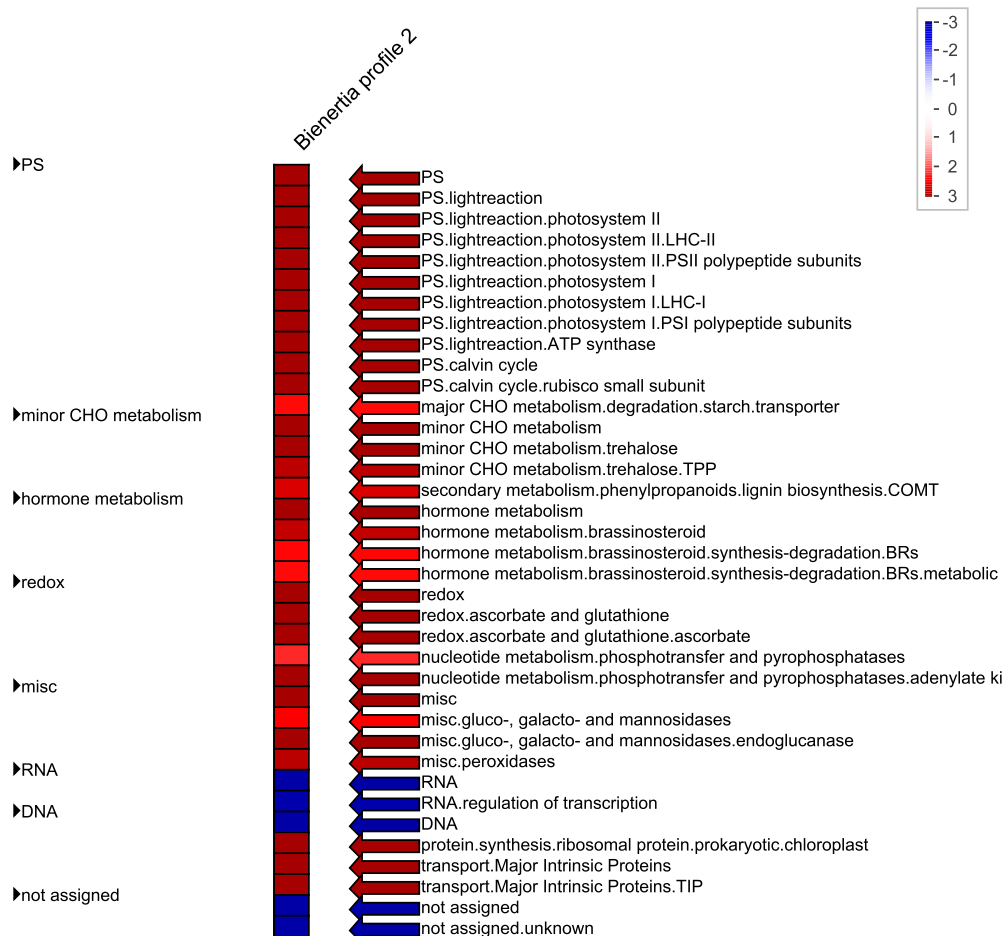
Figure A.2: Enrichment analysis of gene profile 2 from Bienertia.
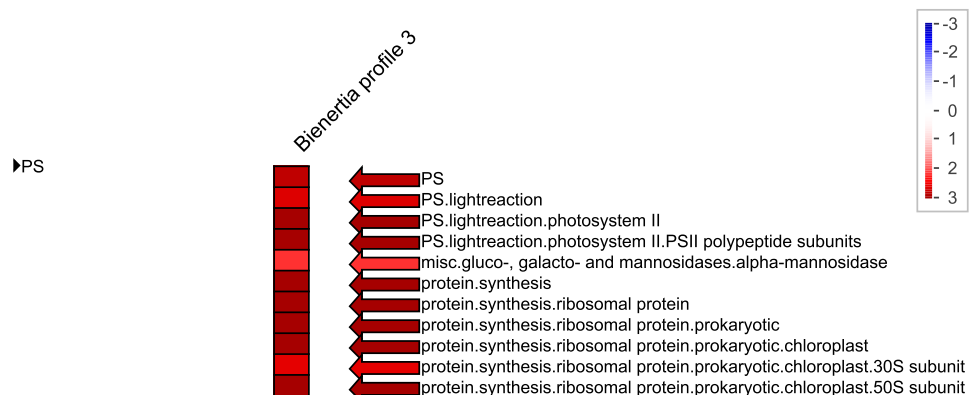


Figure A.3: Enrichment analysis of gene profile 3 from Bienertia.
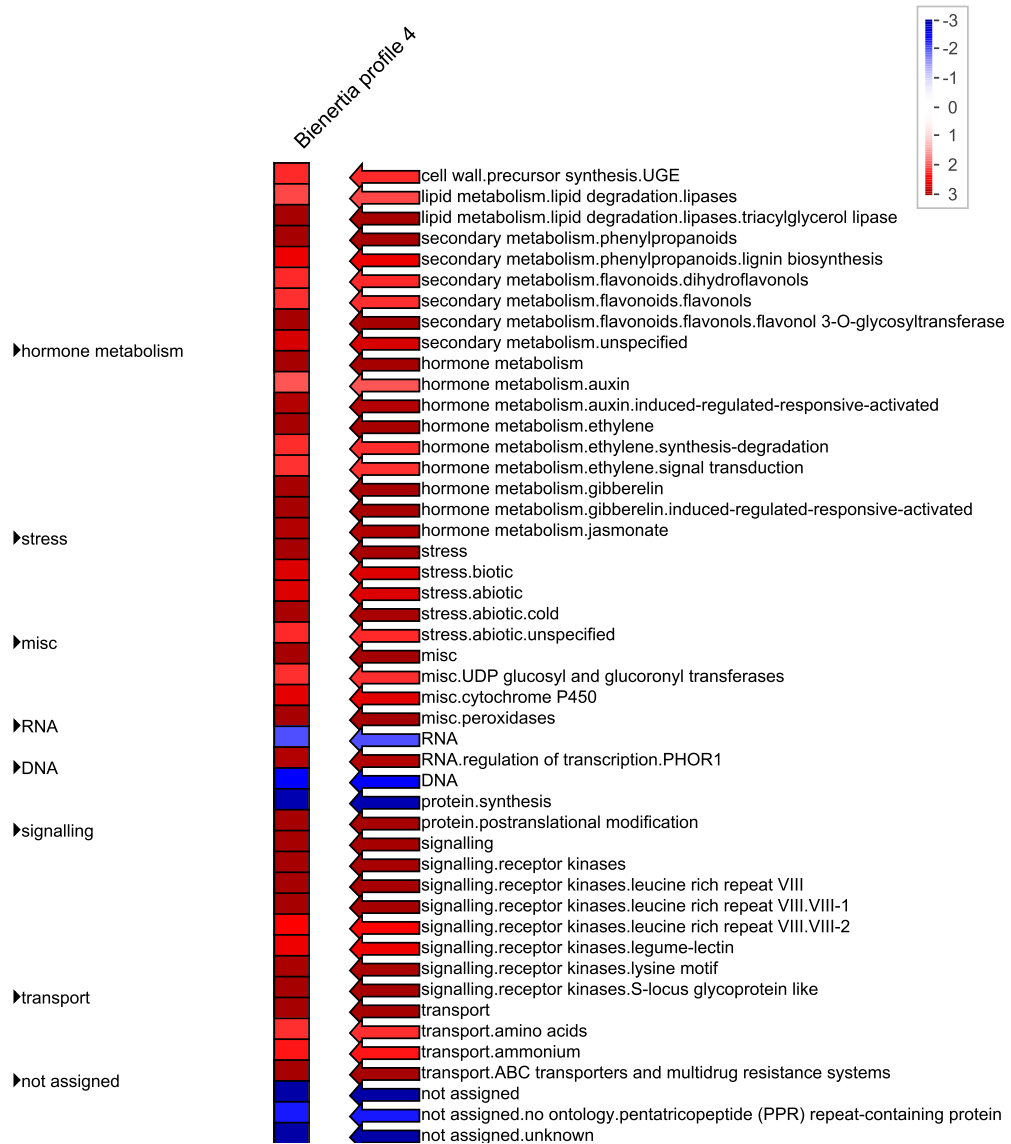
Figure A.4: Enrichment analysis of gene profile 4 from Bienertia.
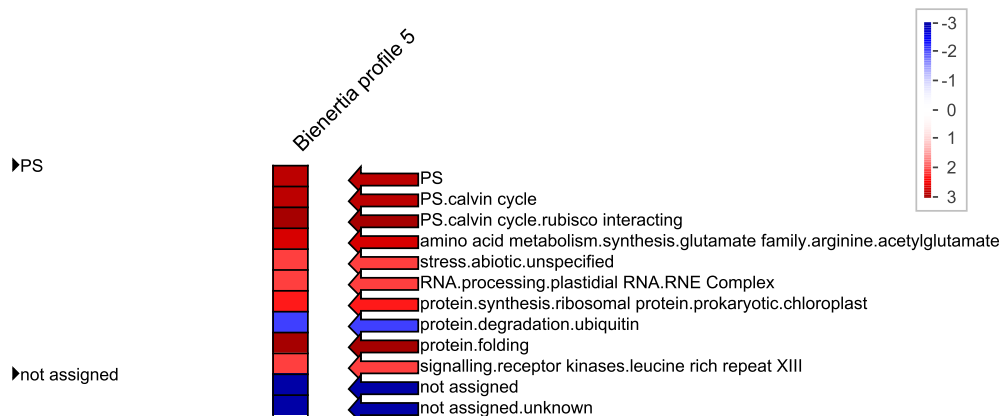
Figure A.5: Enrichment analysis of gene profile 5 from Bienertia.
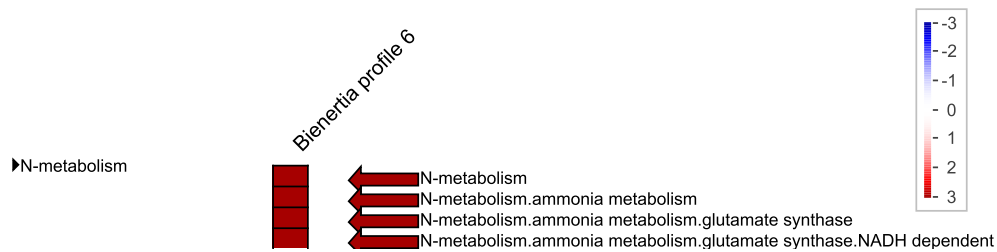


Figure A.6: Enrichment analysis of gene profile 6 from Bienertia.
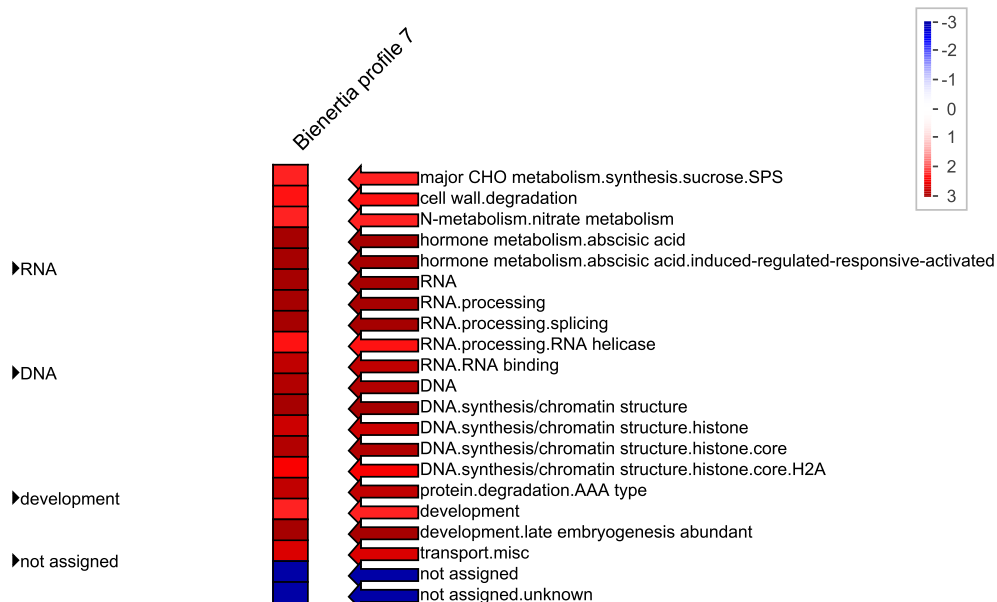


Figure A.7: Enrichment analysis of gene profile 7 from Bienertia.
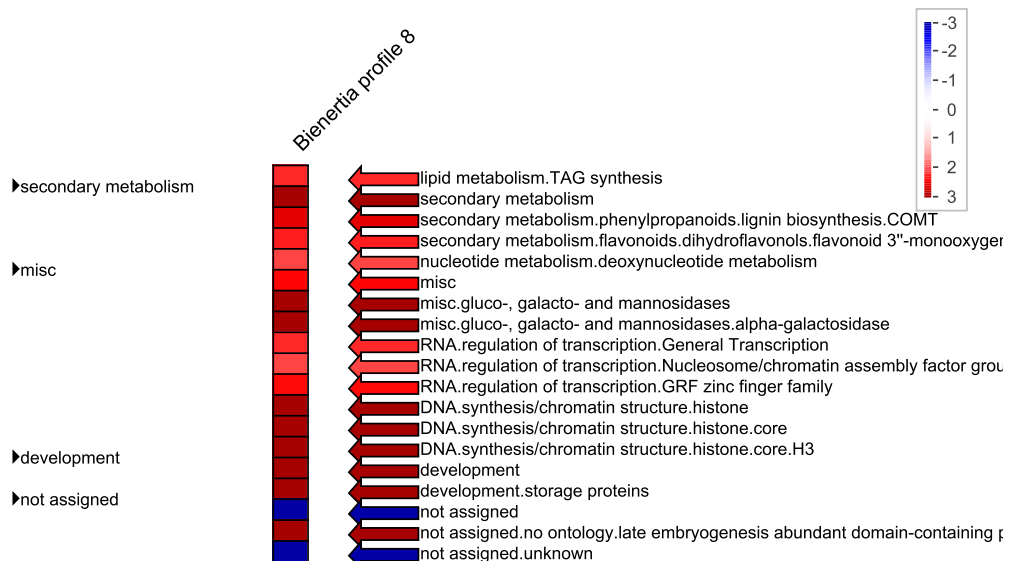
Figure A.8: Enrichment analysis of gene profile 8 from Bienertia.
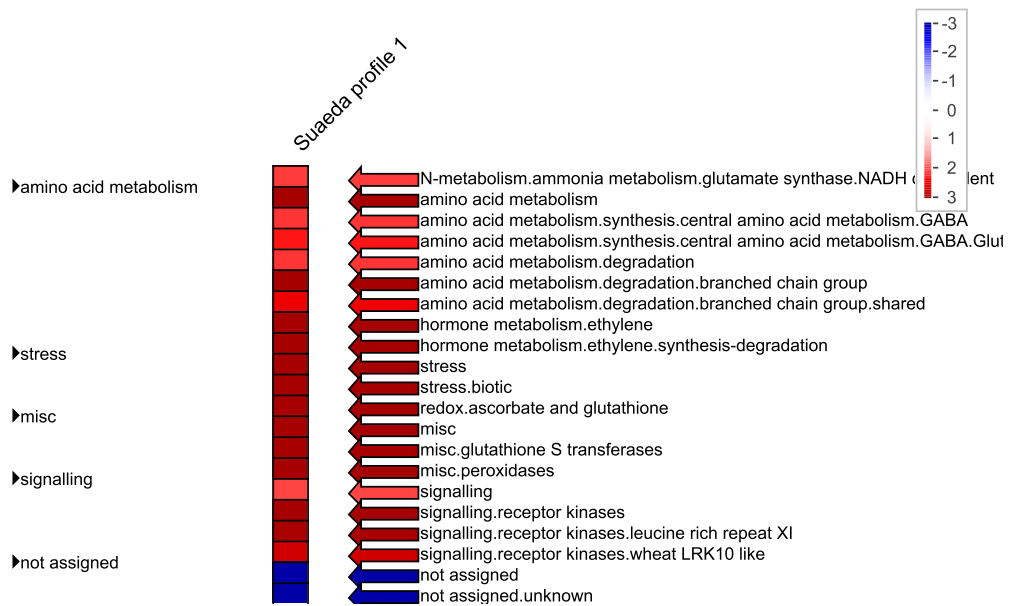


Figure A.9: Enrichment analysis of gene profile 1 from Suaeda.
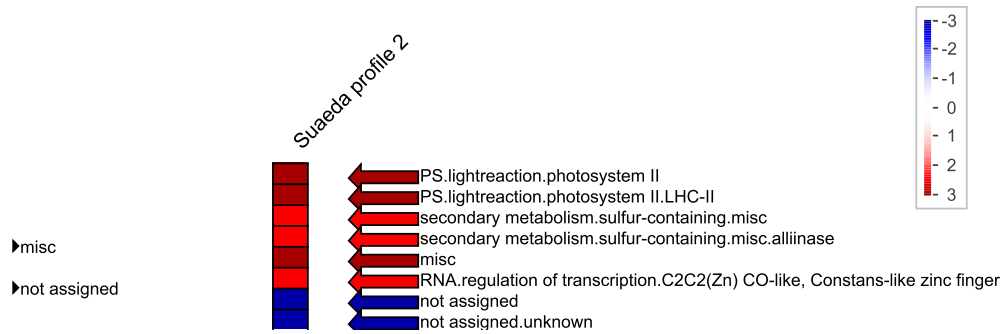
Figure A.10: Enrichment analysis of gene profile 2 from Suaeda.
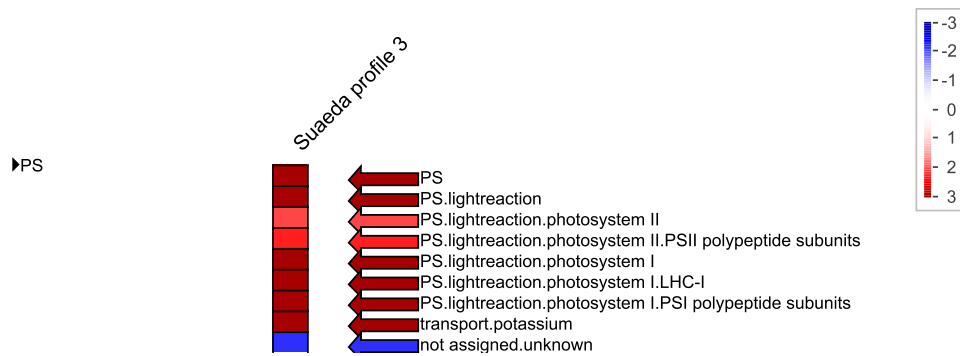


Figure A.11: Enrichment analysis of gene profile 3 from Suaeda.
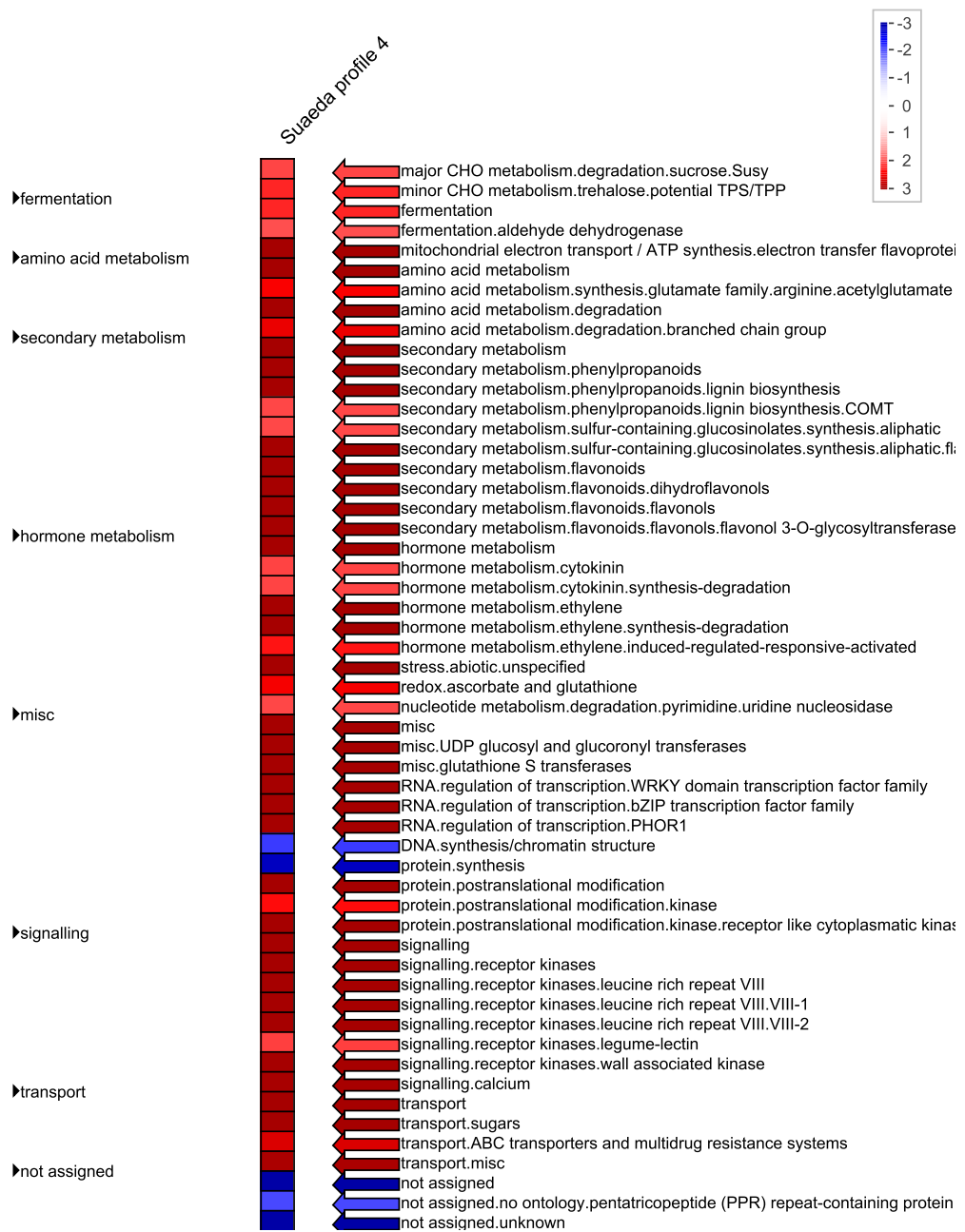
Figure A.12: Enrichment analysis of gene profile 4 from Suaeda.
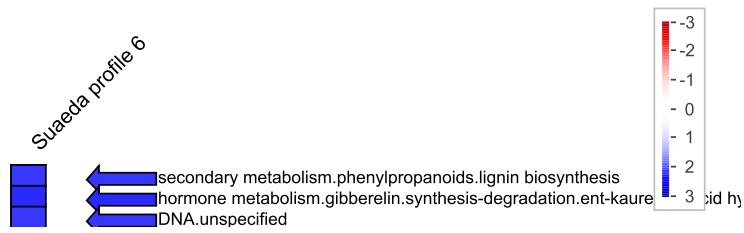
Figure A.13: Enrichment analysis of gene profile 6 from Suaeda.
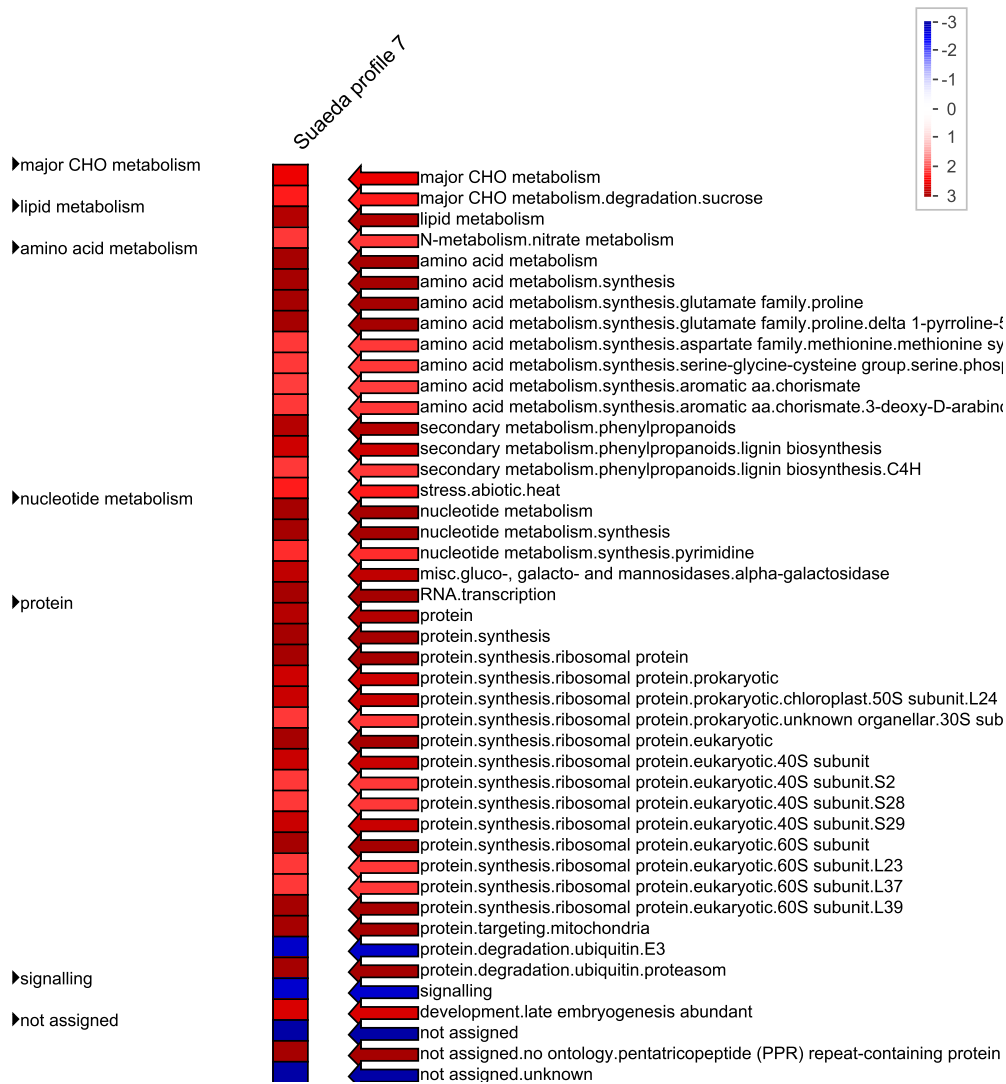


Figure A.14: Enrichment analysis of gene profile 7 from Suaeda.
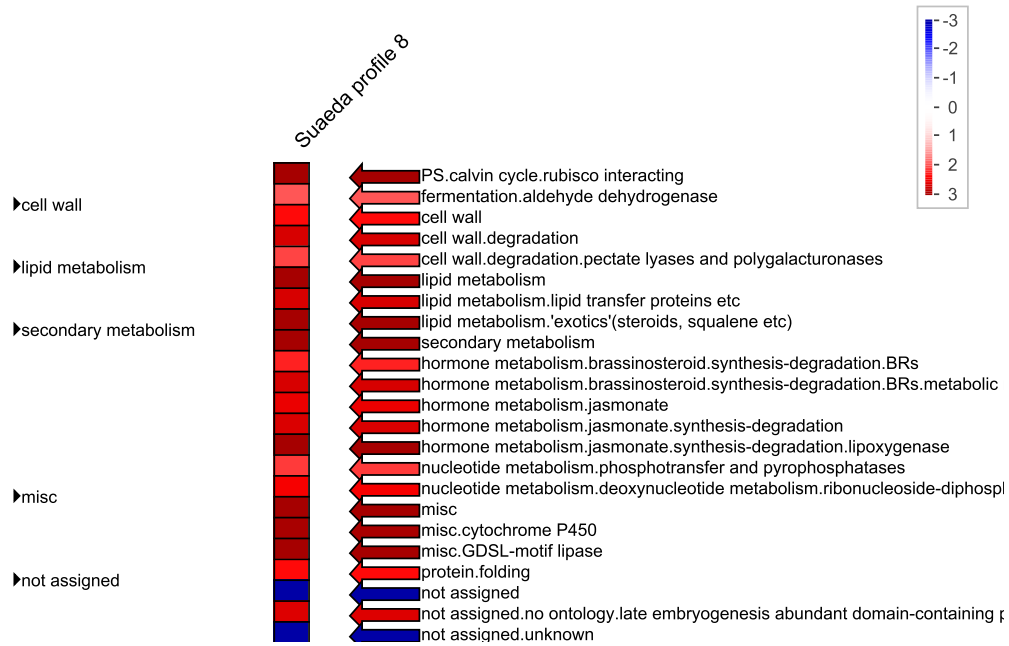
Figure A.15: Enrichment analysis of gene profile 8 from Suaeda.

## A.2 Genes more abundant in Bienertia compared to Suaeda

Table A.1: Abbreviated list of genes more abundant in Bienertia than Suaeda. The first 100 list entries are shown. The complete list is added in electronic form. Only genes with a mean expression > 100 NC in Bienertia are listed. ratio B/S refers to the ratio of mean expression over TP1 to 3 in Bienertia and Suaeda. The predicted function is based on the Mercator annotation of the *B.vulgaris* transcriptome

| gene | mean expression TP1-3 | | ratio B/S | predicted function |
| | Bienertia | Suaeda | | |
|------|-----------|--------|-----------|--------------------|
| Bv_018260_zyau.t1 | 433.32 | 0.00 | na | O-methyltransferase ZRP4 (OMT) |
| Bv1_010170_rxtd.t1 | 125.84 | 0.00 | na | unknown protein |
| Bv2_027480_ehzg.t1 | 133.11 | 0.00 | na | unknown protein |
| Bv4_071390_oznj.t1 | 262.77 | 0.00 | na | Family of unknown function (DUF716) |
| Bv5_099020_yunk.t1 | 217.82 | 0.00 | na | Putrescine N-methyltransferase 4 (PMT 4) |
| Bv6_144370_hxzq.t1 | 100.80 | 0.00 | na | maternal effect embryo arrest 11 (MEE11) |
| Bv7_171060_eoxy.t1 | 182.52 | 0.00 | na | histone deacetylase 8 (HDA08) |
| Bv4_092160_nehd.t1 | 113.97 | 0.00 | na | homolog of yeast autophagy 18 (ATG18) |
| Bv9_213290_wgto.t1 | 369.50 | 0.00 | na | unknown protein |
| Bv9_218100_aipe.t1 | 163.94 | 0.00 | na | unknown protein |
| Bv4_096830_sumq.t1 | 1234.07 | 0.62 | 1983.96 | unknown protein |
| Bv4_074220_juow.t1 | 294.32 | 0.19 | 1548.62 | unknown protein |
| Bv4_096850_htfc.t1 | 944.79 | 0.62 | 1518.90 | unknown protein |
| Bv_004200_dojp.t1 | 569.33 | 0.38 | 1499.51 | Glutaredoxin family protein |
| Bv4_074190_ctkt.t1 | 2064.85 | 1.54 | 1342.57 | unknown protein |

| gene | mean expression TP1-3 | | ratio B/S | predicted function |
|------|-----------|--------|-----------|--------------------|
|      | **Bienertia** | **Suaeda** |  |  |
| Bv4_091130_ekaa.t1 | 428.56 | 0.35 | 1209.58 | Chaperone DnaJ-domain |
| Bv_011480_ohdg.t1 | 246.56 | 0.21 | 1189.14 | ARA6 |
| Bv7_165900_qdtq.t1 | 191.41 | 0.17 | 1110.68 | unknown protein |
| Bv2_032080_gqda.t1 | 2205.60 | 2.08 | 1059.48 | DNAse I-like |
| Bv4_092160_nehd.t2 | 175.14 | 0.19 | 921.51 | homolog of yeast autophagy 18 (ATG18) |
| Bv3_056320_wfdf.t1 | 151.44 | 0.17 | 878.73 | Major facilitator superfamily protein |
| Bv6_144160_zztn.t1 | 165.19 | 0.19 | 869.16 | Putative ion channel DMI-1 |
| Bv8_187590_wmgy.t1 | 471.23 | 0.60 | 779.24 | Auxin-induced in root cultures (AIR9) |
| Bv6_143940_dmte.t1 | 160.93 | 0.21 | 776.15 | beta-1,4-N-acetylglucosaminyltransferase |
| Bv1_015280_udcx.t1 | 3593.14 | 4.89 | 734.33 | cytochrome p450, CYP81F |
| Bv1_012500_wrja.t1 | 149.71 | 0.21 | 722.03 | putative hydroxysteroid dehydrogenase (HSD) |
| Bv3_054050_nmgg.t1 | 143.84 | 0.21 | 693.72 | unknown protein |
| Bv2_025530_okoh.t1 | 3570.27 | 5.24 | 681.79 | unknown protein |
| Bv7_162870_jnmd.t1 | 133.12 | 0.21 | 642.03 | unknown protein |
| Bv6_148000_pfpc.t1 | 979.46 | 1.53 | 638.88 | cytochrome p450, CYP716A |
| Bv7_163120_gqkz.t1 | 260.03 | 0.41 | 627.05 | RING/U-box superfamily protein |
| Bv3_060760_dhyt.t1 | 488.45 | 0.79 | 618.67 | polygalacturonase inhibiting protein 1 (PGIP1) |
| Bv2_037980_csxr.t1 | 339.82 | 0.57 | 596.45 | unknown protein |
| Bv1_021440_jixd.t1 | 463.65 | 0.81 | 570.94 | alpha/beta-Hydrolase |
| Bv6_144160_zztn.t2 | 236.40 | 0.41 | 570.08 | Putative ion channel DMI-1 |
| Bv2_029700_pfgu.t1 | 663.99 | 1.18 | 560.96 | Transmembrane amino acid transporter |
| Bv3_059780_onew.t1 | 293.69 | 0.57 | 515.49 | Concanavalin A-like lectin |
| Bv2_038300_ztiu.t1 | 102.63 | 0.21 | 494.98 | PATATIN-like protein 6 (PLP6) |
| Bv3_060740_mupo.t1 | 982.82 | 2.00 | 492.22 | polygalacturonase inhibiting protein 1 (PGIP1) |
| Bv8_195280_gwfj.t1 | 348.49 | 0.72 | 481.10 | unknown protein |
| Bv_007810_nznd.t1 | 180.87 | 0.38 | 476.37 | Auxin-induced in root cultures (AIR9) |
| Bv8_191040_drte.t1 | 283.97 | 0.60 | 469.57 | Cytochrome P450 |
| Bv8_188140_iwdd.t1 | 924.39 | 2.04 | 453.55 | embryo defective 1241 |
| Bv4_096840_ajge.t1 | 1858.24 | 4.25 | 437.30 | unknown protein |
| Bv5_100780_rgrk.t1 | 1248.36 | 2.91 | 429.51 | Zeamatin precursor |
| Bv2_029680_jmxw.t1 | 401.86 | 1.00 | 401.07 | Transmembrane amino acid transporter |
| Bv3_057670_wahd.t1 | 163.27 | 0.41 | 393.71 | Endochitinase A2 precursor |
| Bv5_124970_jeup.t1 | 162.03 | 0.41 | 390.74 | unknown protein |
| Bv5_119020_xids.t1 | 217.44 | 0.56 | 387.15 | Uridine diphosphate glycosyltransferase |
| Bv9_206920_kkqt.t1 | 582.78 | 1.53 | 382.15 | Microtubule-Associated Stress Protein 1, MASP1 |
| Bv_012410_scnx.t1 | 151.02 | 0.40 | 380.03 | unknown protein |
| Bv_009820_mrec.t1 | 1032.52 | 2.74 | 376.50 | ATP sulfurylase 1 (APS1) |
| Bv5_098990_nnaz.t1 | 237.80 | 0.67 | 357.39 | Cold and drought-regulated protein CORA |
| Bv6_144370_hxzq.t2 | 146.48 | 0.41 | 353.23 | maternal effect embryo arrest 11 (MEE11) |
| Bv_005040_cxok.t1 | 4470.51 | 12.81 | 348.91 | Chlorophyll a-b binding protein CP24 |
| Bv5_121340_cuxq.t1 | 141.75 | 0.41 | 341.84 | PKDM7D |
| Bv5_100060_uwnw.t1 | 439.42 | 1.32 | 333.76 | Basic 7S globulin 2 precursor |
| Bv7_170780_etrj.t1 | 243.71 | 0.76 | 320.94 | decapping 1 (DCP1) |
| Bv9_211420_yrok.t1 | 102.65 | 0.32 | 316.78 | zinc finger (MYND type) family protein |
| Bv3_057620_sqyj.t1 | 570.22 | 1.82 | 314.13 | Endochitinase A2 precursor |
| Bv2_030040_dnyj.t1 | 297.20 | 0.96 | 309.89 | cytochrome p450,CYP81F |
| Bv5_104770_htwy.t2 | 299.47 | 0.97 | 309.78 | cytochrome p450, CYP88A |

| gene | mean expression TP1-3 | | ratio B/S | predicted function |
|------|-----------|--------|-----------|--------------------|
| | **Bienertia** | **Suaeda** | | |
| Bv7_177990_otyh.t1 | 332.30 | 1.08 | 307.93 | camelliol C synthase 1 (CAMS1) |
| Bv3_053580_yyih.t1 | 518.55 | 1.69 | 306.64 | unknown protein |
| Bv4_091140_xstd.t1 | 581.29 | 1.91 | 304.79 | seed imbibition 1 (SIP1) |
| Bv5_104780_kudm.t1 | 2444.28 | 8.20 | 298.24 | cytochrome p450, CYP88A |
| Bv6_145210_uijh.t1 | 355.21 | 1.23 | 289.55 | FAR1-related sequence 11 (FRS11) |
| Bv7_171140_qwks.t1 | 392.42 | 1.38 | 283.99 | LONESOME HIGHWAY (LHW) |
| Bv5_104770_htwy.t1 | 598.29 | 2.26 | 264.34 | cytochrome p450, CYP88A |
| Bv6_130120_uxpp.t1 | 398.18 | 1.52 | 262.15 | Auxin efflux carrier family protein |
| Bv7_163150_tchj.t1 | 158.36 | 0.60 | 261.86 | cyclin p1 (CYCP1) |
| Bv_001660_ufqc.t2 | 155.43 | 0.60 | 257.02 | Zinc finger C-x8-C-x5-C-x3-H type |
| Bv7_158600_gtkn.t1 | 543.77 | 2.19 | 248.80 | IAA-LEUCINE RESISTANT 1 (ILR1) |
| Bv2_034660_tset.t1 | 196.97 | 0.80 | 246.65 | alpha-galactosidase 1 (AGAL1) |
| Bv3_067810_wenn.t1 | 193.89 | 0.79 | 244.08 | Acyl-CoA N-acyltransferase |
| Bv6_137800_eycn.t1 | 226.11 | 0.93 | 242.57 | unknown protein |
| Bv7_171170_cpoh.t1 | 112.45 | 0.50 | 226.27 | plant U-box 24 (PUB24) |
| Bv2_039660_pewg.t1 | 499.72 | 2.32 | 215.56 | serine carboxypeptidase-like 25 (scpl25) |
| Bv6_144020_oadt.t1 | 195.07 | 0.98 | 198.07 | unknown protein |
| Bv5_106490_fawt.t1 | 361.35 | 1.84 | 196.48 | unknown protein |
| Bv6_133120_pfnc.t1 | 222.21 | 1.21 | 183.79 | Autophagy-related protein 13 |
| Bv6_133110_rmmq.t1 | 110.34 | 0.60 | 182.46 | protein kinase |
| Bv7_178040_mfeg.t1 | 2117.53 | 11.72 | 180.67 | Polyamine oxidase precursor |
| Bv7_169120_jeep.t1 | 1190.86 | 6.70 | 177.62 | Pectinesterase-1 precursor |
| Bv7_168340_tgio.t1 | 238.98 | 1.35 | 177.44 | Trypsin family protein with PDZ domain |
| Bv3_060750_yrdi.t1 | 515.45 | 3.00 | 172.09 | polygalacturonase inhibiting protein 1 (PGIP1) |
| Bv7_169320_jtrk.t1 | 1122.23 | 6.81 | 164.83 | Polyamine oxidase precursor |
| Bv5_115750_kuem.t1 | 401.92 | 2.59 | 155.04 | PR (pathogenesis-related) protein |
| Bv7_157540_nnye.t1 | 832.30 | 5.63 | 147.71 | Casbene synthase |
| Bv9_203190_tzns.t1 | 313.77 | 2.16 | 145.20 | glycosyl hydrolase |
| Bv2_042500_ciwn.t1 | 2137.81 | 15.00 | 142.55 | Chalcone synthase 2 |
| Bv_010750_yozs.t1 | 192.18 | 1.36 | 140.88 | transducin/WD40 family |
| Bv5_100820_xcqf.t1 | 3161.36 | 23.10 | 136.86 | protein P21 |
| Bv5_112550_upqh.t1 | 102.43 | 0.75 | 136.19 | PQ-loop repeat family protein |
| Bv8_192220_giyo.t1 | 704.18 | 5.26 | 133.92 | Plant Natriuretic Peptide (PNP) |
| Bv2_023590_fxin.t1 | 177.55 | 1.35 | 131.94 | glycosyl hydrolase |
| Bv1_001880_sums.t1 | 320.35 | 2.49 | 128.61 | nodulin MtN21 /EamA-like transporter |
| Bv4_078920_hyyu.t1 | 365.88 | 2.85 | 128.43 | methyltransferase |
| Bv9_212760_dgaq.t1 | 117.52 | 0.94 | 125.46 | flavin-dependent monooxygenase 1 (FMO1) |
| Bv5_100810_ufoi.t1 | 1085.43 | 8.89 | 122.05 | protein P21 |

# A.3 Comparison of light-induced photosynthetic genes in Bienertia and Arabidopsis

Table A.2: List of light-induced photosynthetic genes in Arabidopsis and Bienertia. The fold-change for Arabidopsis was calculated from microarray data of 2 and 5 day old seedlings grown in darkness and light. The foldchange for Bienertia was calculated with DESeq2 from 5 day old cotyledons exposed to light for 48 hours. Mean BS [NC] refers to the mean expression over TP1 to 3 in Bienertia. FC: fold change; AT: Arabidopsis; BV: *B. vulgaris*; BS: Bienertia

| AT gene | BV gene | gene name | FC AT | FC BS | mean BS [NC] |
|---------|---------|-----------|-------|-------|--------------|
| AT4G26530 | Bv4_091040_cyuu.t1 | FBA5 | 67.50 | 9.24 | 714.54 |
| AT3G27690 | Bv5_122230_sssd.t1 | LHCB2.4 | 30.02 | 1.89 | 479.75 |
| AT2G39470 | Bv2_037200_dtwf.t1 | PNSL1 | 28.14 | 3.38 | 479.65 |
| AT1G70760 | Bv9_224960_msjj.t1 | ndhL | 26.66 | 3.02 | 1126.45 |
| AT3G16250 | Bv9_221220_moua.t1 | PNSB3 | 25.59 | 2.61 | 125.40 |
| AT1G14150 | Bv6_153560_keep.t1 | PNSL2 | 19.33 | 2.13 | 696.82 |
| AT3G54050 | Bv5_111310_qxks.t1 | FBP | 18.87 | 1.85 | 1982.00 |
| AT4G33010 | Bv_012000_yknj.t1 | GLDP1 | 15.61 | 5.41 | 10815.05 |
| AT3G55800 | Bv2_024290_rmpf.t1 | SBPASE | 14.71 | 1.31 | 14347.15 |
| AT4G28660 | Bv7_166610_cumn.t1 | PSB28 | 14.28 | 1.77 | 538.54 |
| AT5G36700 | Bv4_074740_miaa.t1 | PGLP1B | 13.84 | 4.55 | 392.06 |
| AT5G58260 | Bv1_007840_ynwt.t1 | ndhN | 13.07 | 2.42 | 1407.29 |
| AT1G70580 | Bv6_148110_nuir.t1 | GGAT2 | 12.67 | 3.48 | 3366.74 |
| AT1G19150 | Bv8_190730_noee.t1 | LHCA6 | 12.14 | 1.94 | 1817.49 |
| AT4G09650 | Bv6_150240_akyy.t1 | ATPD | 11.92 | 3.33 | 3993.73 |
| AT4G22890 | Bv5_101470_qztq.t1 | PGRL1A | 11.47 | 3.93 | 482.13 |
| AT3G01440 | Bv4_092210_gfug.t1 | PNSL3 | 11.31 | 1.37 | 524.19 |
| AT2G01590 | Bv1_011040_umth.t1 | CRR3 | 10.58 | 3.83 | 276.81 |
| AT1G77090 | Bv1_010720_weiu.t1 | PPD4 | 10.42 | 1.17 | 52.41 |
| AT1G42970 | Bv1_014260_waoc.t1 | GAPB | 10.40 | 1.71 | 16351.37 |
| AT1G32470 | Bv5_106360_ipey.t1 | GDH3 | 10.19 | 1.98 | 1885.81 |
| AT2G21330 | Bv2_039150_dmqz.t1 | FBA1 | 9.60 | 4.81 | 2.91 |
| AT1G52230 | Bv9_221000_fyok.t1 | PSAH2 | 8.44 | 1.62 | 10505.05 |
| AT1G68830 | Bv6_155590_ugjw.t1 | STN7 | 8.39 | 7.17 | 199.37 |
| AT3G50820 | Bv4_072250_ynjs.t1 | PSBO2 | 7.61 | 1.19 | 32670.23 |
| AT1G44575 | Bv1_020410_dkjg.t1 | PSBS | 7.33 | 1.06 | 5793.15 |
| AT4G05180 | Bv8_186680_fswh.t1 | PSBQ2 | 7.24 | 1.49 | 9491.68 |
| AT5G36120 | Bv9_203300_uknn.t1 | CCB3 | 7.23 | 1.22 | 70.37 |
| AT1G30380 | Bv7_176760_ahon.t1 | PSAK | 7.11 | 1.66 | 10292.99 |
| AT2G31040 | Bv2_046130_pqko.t1 | CGL160 | 7.00 | 1.47 | 334.26 |
| AT5G66190 | Bv6_130550_saor.t1 | LFNR1 | 6.95 | 1.69 | 4287.55 |
| AT3G21055 | Bv2_029260_fpsj.t1 | PSBT | 6.86 | 1.05 | 1507.42 |
| AT1G32060 | Bv5_107370_rsxd.t1 | PRK | 6.84 | 1.35 | 8317.11 |
| AT1G56190 | Bv2_038200_ezcp.t1 | CPGK2 | 6.83 | 1.24 | 15173.14 |
| AT3G13470 | Bv3_055040_usdd.t1 | CPN60B2 | 6.78 | 0.53 | 7367.51 |
| AT1G45474 | Bv4_089040_okwg.t1 | LHCA5 | 6.71 | 3.55 | 1390.97 |
| AT1G26230 | Bv7_178420_xsah.t1 | CPN60B4 | 6.53 | 1.77 | 378.02 |
| ATCG00270 | Bv5_110350_pzrc.t1 | PSBD | 6.42 | 1.12 | 36070.09 |

| AT gene | BV gene | gene name | FC AT | FC BS | mean BS [NC] |
|---------|---------|-----------|-------|-------|--------------|
| AT2G05620 | Bv3_048340_odem.t1 | PGR5 | 6.38 | 9.63 | 85.40 |
| AT1G76450 | Bv1_013940_wtwf.t1 | PPD3 | 6.29 | 3.13 | 218.68 |
| AT3G55330 | Bv5_123870_tgrg.t1 | PPL1 | 6.25 | 1.57 | 427.42 |
| ATCG00280 | Bv7_180220_zaaj.t1 | PSBC | 6.16 | 1.00 | 8492.14 |
| AT4G15510 | Bv4_080040_qyih.t1 | PPD1 | 6.06 | 0.84 | 378.53 |
| AT3G04790 | Bv4_083630_jofx.t1 | RPI3 | 6.03 | 2.37 | 4902.55 |
| AT1G76570 | Bv1_013790_jjwp.t1 | LHCB7 | 5.90 | 2.86 | 35.33 |
| AT4G32590 | Bv2_043570_rmou.t1 | 2Fe-2S ferredoxin | 5.87 | 0.82 | 176.00 |
| AT1G64770 | Bv3_056240_yegx.t1 | PNSB2 | 5.79 | 2.44 | 2966.07 |
| AT3G56650 | Bv3_063420_jqsz.t1 | PPD6 | 5.75 | 1.59 | 152.44 |
| AT1G08380 | Bv4_087000_zjqt.t1 | PSAO | 5.74 | 2.20 | 782.52 |
| AT5G02120 | Bv2_040200_zwdc.t1 | HLIP | 5.62 | 1.52 | 126.15 |
| AT1G11860 | Bv1_008350_kfrn.t1 | GDCST | 5.43 | 3.11 | 66.72 |
| AT1G10960 | Bv8_196730_uzpt.t1 | FD1 | 5.28 | 2.19 | 1026.87 |
| AT1G55490 | Bv9_221870_ruqu.t1 | CPN60B1 | 5.20 | 0.53 | 5152.35 |
| AT1G73110 | Bv8_189100_paaj.t1 | | 5.12 | 1.22 | 734.65 |
| AT3G15840 | Bv9_219510_hdyh.t1 | PIFI | 5.07 | 3.61 | 743.81 |
| AT1G15820 | Bv_005040_cxok.t1 | LHCB6 | 4.72 | 2.10 | 4470.51 |
| AT1G12900 | Bv4_094590_jftx.t1 | GAPA2 | 4.50 | 1.67 | 24061.18 |
| AT4G02630 | Bv1_015860_hynr.t1 | | 4.44 | 1.31 | 16.42 |
| AT4G17360 | Bv5_104510_raea.t1 | PURU2 | 4.31 | 1.62 | 66.60 |
| AT2G28000 | Bv9_205310_zyqz.t1 | CPN60A1 | 4.29 | 0.43 | 13665.35 |
| AT1G67740 | Bv6_147470_qqmp.t1 | PSBY | 4.22 | 1.76 | 6770.85 |
| AT5G38660 | Bv2_026500_swcr.t1 | APE1 | 4.20 | 1.86 | 376.73 |
| AT3G12780 | Bv7_161710_aszc.t1 | PGK1 | 4.18 | 1.11 | 39.32 |
| AT3G54890 | Bv2_047600_uhic.t1 | LHCA1 | 4.15 | 1.72 | 12495.61 |
| AT4G38970 | Bv6_129880_fskg.t1 | FBA2 | 3.96 | 3.06 | 2639.38 |
| AT5G11450 | Bv2_043690_zctz.t1 | PPD5 | 3.86 | 0.46 | 619.00 |
| AT1G74880 | Bv6_132550_qnan.t1 | ndhO | 3.64 | 2.66 | 360.37 |
| AT4G04640 | Bv8_186220_smue.t1 | ATPC1 | 3.56 | 1.33 | 7269.27 |
| AT5G23120 | Bv9_203050_uiqh.t1 | HCF136 | 3.54 | 1.56 | 1616.01 |
| AT2G28605 | Bv4_078670_puyq.t1 | PPD2 | 3.52 | 1.41 | 73.54 |
| AT1G68010 | Bv9_213980_zwen.t1 | HPR | 3.48 | 2.35 | 1980.62 |
| AT4G02770 | Bv6_139980_qgth.t1 | PSAD1 | 3.43 | 1.45 | 17408.49 |
| AT1G20340 | Bv_004160_hgjn.t1 | DRT112 | 3.43 | 1.18 | 7716.66 |
| AT1G34000 | Bv1_012160_icjc.t1 | OHP2 | 3.42 | 1.94 | 258.66 |
| AT3G14420 | Bv4_094290_jgpp.t1 | GLO1 | 3.35 | 3.82 | 6870.27 |
| AT1G51400 | Bv4_072800_azzt.t1 | PS II 5 kD | 3.11 | 1.62 | 944.10 |
| AT5G51545 | Bv3_055830_swgf.t1 | LPA2 | 3.10 | 1.55 | 25.05 |
| AT1G15140 | Bv6_142970_whmk.t1 | PS II 5 kD | 3.01 | 1.71 | 250.50 |
| AT1G05385 | Bv2_032950_jtme.t1 | PSB27-2 | 2.99 | 0.83 | 43.03 |
| AT5G54270 | Bv4_093140_qsnk.t1 | LHCB3 | 2.96 | 6.78 | 1876.01 |
| AT2G39730 | Bv2_025300_tzou.t1 | RCA | 2.96 | 2.67 | 63931.21 |
| AT5G18820 | Bv6_156500_ymir.t1 | CPN60A2 | 2.94 | 0.73 | 106.04 |
| AT3G60750 | Bv3_055740_ueqe.t1 | TKL-1 | 2.91 | 1.12 | 54432.06 |
| AT2G05100 | Bv9_225010_yekr.t1 | LHCB2.1 | 2.88 | 1.90 | 63426.29 |
| AT5G64040 | Bv6_132600_isrd.t1 | PSAN | 2.84 | 1.71 | 2599.44 |
| AT1G14030 | Bv6_153430_xikk.t1 | LSMT-L | 2.81 | 0.80 | 704.38 |

| AT gene | BV gene | gene name | FC AT | FC BS | mean BS [NC] |
|---|---|---|---|---|---|
| AT2G46820 | Bv1_002080_grta.t1 | CURT1B | 2.78 | 4.35 | 1138.70 |
| AT5G64380 | Bv9_221420_ohea.t1 | | 2.74 | 1.56 | 635.35 |
| AT3G23990 | Bv_042320_jjwq.t1 | CPN60 | 2.70 | 0.41 | 1.87 |
| AT4G20130 | Bv3_061710_eiwy.t1 | PTAC14 | 2.69 | 0.82 | 589.33 |
| ATCG00130 | Bv4_094260_gpgr.t1 | ATPF | 2.66 | 0.91 | 28189.81 |
| AT1G06680 | Bv6_132120_uwxc.t1 | PSBP1 | 2.46 | 1.28 | 27729.25 |
| AT2G20260 | Bv7_166490_qiwa.t1 | PSAE2 | 2.45 | 1.58 | 5580.99 |
| ATCG00300 | Bv7_180210_mxfx.t1 | PSBZ | 2.40 | 0.68 | 2738.92 |
| AT2G06520 | Bv8_192880_cyec.t1 | PSBX | 2.40 | 1.59 | 7394.00 |
| AT4G32260 | Bv_012930_esja.t1 | PDE334 | 2.39 | 2.67 | 8381.88 |
| AT1G03600 | Bv7_173090_nqux.t1 | PSB27-1 | 2.34 | 1.32 | 892.37 |
| AT4G03280 | Bv9_206750_aefc.t1 | PETC | 2.21 | 3.41 | 6954.89 |
| ATCG00350 | Bv1_010710_uohn.t1 | PSAA | 2.20 | 1.01 | 24495.56 |
| AT2G30570 | Bv5_120370_zigj.t1 | PSBW | 2.18 | 1.68 | 7164.89 |
| AT5G61410 | Bv5_114350_cnci.t1 | RPE | 2.17 | 1.81 | 4919.23 |
| ATCG00140 | Bv3_057950_wahj.t1 | ATPH | 2.12 | 0.97 | 9357.84 |
| AT4G10340 | Bv1_004060_xhfp.t1 | LHCB5 | 2.00 | 1.44 | 57146.30 |
| ATCG00150 | Bv_032280_njuh.t1 | ATPI | 1.98 | 0.94 | 29995.33 |
| AT3G09150 | Bv3_060200_ezrp.t1 | HY2 | 1.98 | 1.08 | 22.90 |
| AT4G12800 | Bv4_075220_usks.t1 | PSAL | 1.95 | 1.38 | 31579.65 |
| ATCG00680 | Bv1_008900_pnio.t1 | PSBB | 1.92 | 0.66 | 40526.80 |
| AT1G61520 | Bv1_021010_dyhz.t1 | LHCA3 | 1.86 | 2.13 | 10570.31 |
| ATCG00120 | Bv8_201580_ruit.t1 | ATPA | 1.85 | 0.83 | 12820.39 |
| AT1G31330 | Bv9_214300_nfsx.t1 | PSAF | 1.72 | 1.31 | 36431.41 |
| AT4G32520 | Bv2_033270_kaqs.t1 | SHM3 | 1.72 | 0.01 | 1.63 |
| AT3G47470 | Bv5_114540_xkts.t1 | LHCA4 | 1.71 | 4.68 | 7159.58 |
| AT5G03690 | Bv7_160160_nqkh.t1 | FBA4 | 1.69 | 0.92 | 23169.60 |
| AT5G01530 | Bv2_043310_ppkn.t1 | LHCB4.1 | 1.67 | 1.86 | 7355.65 |
| AT2G21170 | Bv9_214650_tday.t1 | TIM | 1.66 | 1.08 | 3757.79 |
| AT1G79870 | Bv9_220970_zhqq.t1 | HPR2 | 1.66 | 1.62 | 39.43 |
| AT3G61470 | Bv2_033010_pnos.t1 | LHCA2 | 1.65 | 1.55 | 14121.44 |
| AT1G55670 | Bv3_065490_nptx.t1 | PSAG | 1.62 | 2.22 | 6078.31 |
| ATCG00340 | Bv7_180200_pzut.t1 | PSAB | 1.60 | 0.97 | 45049.83 |
| AT2G45290 | Bv3_070720_qekf.t1 | TKL-2 | 1.59 | 1.16 | 1176.84 |
| AT5G47760 | Bv8_184280_guso.t1 | PGLP2 | 1.58 | 1.10 | 337.85 |
| AT1G60950 | Bv2_024300_ciwa.t1 | FD2 | 1.58 | 2.55 | 2933.34 |
| AT4G00895 | Bv1_002600_euia.t1 | ATPF1 | 1.57 | 1.65 | 196.04 |
| AT1G29930 | Bv7_175110_jsio.t1 | LHCB1.3 | 1.55 | 2.48 | 93485.96 |
| AT1G79550 | Bv2_038140_gczg.t1 | PGK | 1.51 | 1.00 | 14474.21 |
| AT2G26500 | Bv1_012920_jdxy.t1 | petM | 1.49 | 1.68 | 2618.82 |
| AT1G79040 | Bv4_096020_ahue.t1 | PSBR | 1.46 | 2.10 | 12431.56 |
| ATCG01110 | Bv5_124770_udnj.t1 | NDHH | 1.45 | 1.02 | 17813.74 |
| AT1G01540 | Bv1_004150_qicq.t1 | | 1.40 | 1.06 | 392.61 |
| AT4G32360 | Bv8_196850_wtrh.t1 | MFDR | 1.40 | 1.14 | 107.47 |
| ATCG00430 | Bv_010270_ngxg.t1 | NDHK | 1.38 | 1.48 | 18467.43 |
| ATCG00470 | Bv_030820_rwta.t1 | ATPE | 1.37 | 0.85 | 7833.03 |
| AT2G35120 | Bv2_044720_cgkg.t1 | GDH2 | 1.27 | 0.77 | 481.39 |
| ATCG00580 | Bv2_037630_qtrs.t1 | PSBE | 1.26 | 0.68 | 6320.71 |

Table A.2: (continued)

| AT gene | BV gene | gene name | FC AT | FC BS | mean BS [NC] |
|---------|---------|-----------|-------|-------|--------------|
| AT4G22260 | Bv8_185240_zzoj.t1 | AOX4 | 1.24 | 1.18 | 639.14 |
| ATCG01250 | Bv5_108370_jyyy.t1 | NDHB.2 | 1.23 | 0.49 | 3883.67 |
| ATCG00420 | Bv7_165970_dpqj.t1 | NDHJ | 1.22 | 1.43 | 24483.28 |
| AT2G13360 | Bv4_073470_iswc.t1 | AGT1 | 1.20 | 8.35 | 3177.80 |
| AT1G67090 | Bv2_026840_jycs.t1 | RBCS1A | 1.20 | 1.48 | 16487.87 |
| ATCG00540 | Bv3_055950_cher.t1 | PETA | 1.19 | 1.06 | 47483.80 |
| ATCG00480 | Bv3_069750_thqf.t1 | ATPB | 1.19 | 0.83 | 19857.45 |
| AT3G21740 | Bv2_028630_mzsi.t1 | APO4 | 1.16 | 0.41 | 160.59 |
| ATCG01080 | Bv7_180230_hhnw.t1 | NDHG | 1.15 | 0.86 | 13324.03 |
| AT3G14130 | Bv7_169640_fmtj.t1 | GLO4 | 1.12 | 0.86 | 146.31 |
| ATCG00080 | Bv_017090_ztdm.t1 | PSBI | 1.12 | 0.82 | 32171.31 |
| AT2G47400 | Bv5_121790_sfch.t1 | CP12-1 | 1.10 | 2.28 | 4039.26 |
| ATCG00490 | Bv3_057640_epim.t1 | RBCL | 1.06 | 1.00 | 569.96 |
| AT1G12550 | Bv5_101500_mrdf.t1 | HPR3 | 1.06 | 0.52 | 136.65 |
| ATCG00440 | Bv7_180260_puic.t1 | NDHC | 1.05 | 1.37 | 25723.06 |
| ATCG00570 | Bv_027540_stou.t1 | PSBF | 1.03 | 0.74 | 4465.12 |
| AT2G45630 | Bv_011540_wfpe.t1 | | 1.00 | 0.79 | 43.81 |
| ATCG01010 | Bv1_006390_hkjf.t1 | NDHF | 0.98 | 0.83 | 837.47 |
| ATCG00720 | Bv1_008880_kiqi.t1 | PETB | 0.93 | 0.65 | 11870.35 |
| ATCG01090 | Bv5_124780_ejzg.t1 | NDHI | 0.92 | 0.97 | 22405.92 |
| AT1G02180 | Bv1_017240_tjaz.t1 | | 0.91 | 1.75 | 7.00 |
| ATCG00020 | Bv_037340_xcws.t1 | PSBA | 0.88 | 2.17 | 2869.87 |
| ATCG01050 | Bv6_156290_gykp.t1 | NDHD | 0.75 | 0.77 | 2269.52 |
| AT2G01140 | Bv4_078150_cgxu.t1 | FBA3 | 0.73 | 1.34 | 1169.78 |
| ATCG00630 | Bv7_180630_wfsf.t1 | PSAJ | 0.71 | 0.55 | 323.80 |
| AT5G07950 | Bv4_095820_tdqs.t1 | | 0.70 | 0.26 | 2.69 |
| AT1G80380 | Bv9_220360_xogt.t1 | GLYK | 0.65 | 2.24 | 72.74 |
| ATCG01060 | Bv7_168100_tunx.t1 | PSAC | 0.49 | 0.66 | 7941.94 |
| AT2G25220 | Bv7_166010_ejnm.t1 | | 0.41 | 1.47 | 171.39 |

# Danksagung

Mein größter Dank gilt Dr. Sascha Offermann für die Möglichkeit in seiner Arbeitsgruppe zu promovieren. Du hast mir immer sehr viel Freiraum bei meiner Arbeit gelassen, was ich sehr schätze. Wenn ich mich dann in diverse Analysemethoden verrannt habe, haben Gespräche mit dir mich wieder auf die relevanten Dinge (oder wenigstens auf lösbare Probleme) gelenkt.

Prof. Dr. Helge Küster und Prof. Dr. Jutta Papenbrock danke ich für die Übernahme meines Koreferats und des Promotionsvorsitzes.

Diana, Matthias, Steffi und Philipp: euch danke ich für die sehr schöne und entspannte Stimmung in Büro und Labor. Fünf Leute auf 15 qm kann eigentlich auf Dauer nicht gut gehen, ging aber doch. Lustige Tiervideos und Gespräche, die des Öfteren im Absurden landeten, helfen da natürlich auch.

Der AG Peterhänsel danke ich für die nette Arbeitsatmosphäre, die Hilfsbereitschaft bei sämtlichen Fragestellungen und die Bierfreitage.

Liebe Familie, danke dafür, dass ihr mir ein Studium ermöglicht habt und dass ihr meine Tendenz, ein paar Schlenker in den Lebenslauf einzubauen, immer unterstützt. Ein großer Dank auch meinen Freunden, ganz besonders Silke, für den Rückhalt wenns mal nicht so lief. Ohne euch hätte das alles nur halb so viel Spaß gemacht.

# Lebenslauf

| | |
|---|---|
| Name | Lisa Hagenau |
| Date of Birth | 07.02.1983 |
| Place of Birth | Düren |
| Nationality | Deutsch |

**Publikationen**

Matthias Erlinghäuser, Lisa Hagenau, Diana Wimmer, Sascha Offermann. **Development, subcellular positioning and selective protein accumulation in the dimorphic chloroplasts of single-cell $C_4$ species**. *Current Opinion in Plant Biology*, 31 (2016) pp. 76-82

**Akademische Ausbildung**

| | |
|---|---|
| seit März 2012 | wissenschaftliche Mitarbeiterin/Promotion in der AG Offermann, Institut für Botanik, Leibniz Universität Hannover |
| 2003 - 2010 | Studium der Biologie (Diplom) an der Universität Konstanz |
| | Diplomarbeit bei Prof. Iwona Adamska "Characterisation of stress-enhanced protein 3b in Arabidopsis thaliana using Arabidopsis thaliana mutants" |

**Schulausbildung**

| | |
|---|---|
| 1993 - 2002 | Rurtalgymnasium Düren, Abitur |
| 1989 - 1993 | Grundschule Gürzenich |