Metabolic, regulatory and structural adaptation to C_4 photosynthesis in the genus *Cleome*

Inaugural-Dissertation

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

vorgelegt von

Manuel Sommer

Düsseldorf, Februar 2015

aus dem Institut für Biochemie der Pflanzen
der Heinrich-Heine-Universität Düsseldorf

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

Referent: Prof. Dr. Andreas P. M. Weber
Koreferent: Prof. Dr. Peter Westhoff

Tag der mündlichen Prüfung: 23.03.2015

Erklärung:

Ich versichere an Eides statt, dass ich die vorliegende Dissertation eigenständig und ohne unerlaubte Hilfe unter Beachtung der “Grundsätze zur Sicherung guter wissenschaftlicher Praxis“ an der Heinrich-Heine-Universität Düsseldorf angefertigt habe. Die Dissertation habe ich in dieser oder in ähnlicher Form noch bei keinem anderen Institut eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Düsseldorf, den 18.02.2015

_____

Manuel Sommer

meiner Familie

Contents

Abbreviations **2**

I.1 Summary **4**

I.2 Zusammenfassung **6**

II. Introduction **8**

III. Manuscripts **17**

III.1 Manuscript 1

The dicotyledonous NAD malic enzyme C_4 plant *Cleome gynandra* displays age-dependent plasticity of C_4 decarboxylation biochemistry **17**

III.2 Manuscript 2

Comparative Transcriptome Atlases Reveal Altered Gene Expression Modules between Two Cleomaceae C_3 and C_4 Plant Species **27**

III.3 Manuscript-draft 3

Glk1 function is conserved between *Cleome gynandra* and *Arabidopsis thaliana* as indicated by partial rescue of the broadly de-regulated transcriptome of the *glk*1 *glk*2 double mutant **80**

III.4 Manuscript-draft 4

A specific, cholesterol-binding Lipid Transfer Protein is highly abundant in the C_4 plant *Cleome gynandra* **114**

III.5 Manuscript-draft 5

Plasticity of C_4 photosynthesis in the amphibious sedge *Eleocharis retroflexa* **133**

IV. Acknowledgements **181**

V. Attachments

CD with large supplemental datasets of all manuscripts ..... **last page**

Abbreviations

BS	Bundle sheath
Co1.1	*Cleome gynandra* Golden2-like 1 complementing line 1.1
Co2.1	*Cleome gynandra* Golden2-like 1 complementing line 2.1
D1	Photosystem II subunit A
DGDG	Digalactosyl diacylglycerol
DNA	Desoxyribonucleid acid
Fv/Fm	Maximum quantum yield of photosystem II
GC/MS	Gas chromatography coupled to mass spectrometry
GCT box	Golden2-like / C-terminal box
GFP	Green fluorescent protein
GLK	Golden2-like
GUS	ß-Glucuronidase
HIS	6 x Histidine peptide
HRP	Horseradish peroxidase
LHC	Light harvesting complex
LTP	Lipid transfer protein
M	Mesophyll
MGDG	Monogalactosyl diacylglycerol
NADH	Nicotinamide adenine dinucleotide (reduced)
NADH-DH	Nicotinamide adenine dinucleotide dehydrogenase
NADME	Nicotinamide adenine dinucleotide-dependent malic enzyme
NPQ	Non-photochemical quenching
nsLTP	Non-specific lipid transfer protein
OEC	Oxygen evolving complex
PA	Phosphatidic acid
PC	Phosphatidylcholine
PCA	Principal component analysis
PCR	Polymerase chain reaction
PE	Phosphatidylethanolamine
PEPCK	Phosphoenolpyruvate carboxykinase
PG	Phosphatidylglycerol
PI	Phosphatidylinositol

PsbQ2	Photosystem II subunit Q2
PSI	Photosystem I
PSII	Photosystem II
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RPKM	Reads per kilobase and million
RubisCO	Ribulose-1,5-bisphosphate carboxylase oxygenase
TLC	Thin layer chromatography
YFP	Yellow fluorescent protein

I.1 Summary

C_4 photosynthesis is a very productive trait that evolved more than 66 times independently in autotrophic organisms. C_4 plants split up photosynthesis between two distinct cell types. The biochemistry of the central C_4 pathway interconnecting these cells and the leaf ultrastructure are well studied. In contrast, comparatively little is known about regulatory processes and development of the C_4 trait. Also, the molecular nature of the structural adaptation to C_4 was not described yet. This thesis aims to contribute to a better understanding of a plant's global adaptations to the C_4 trait.

As a model species for the analysis of the C_4 trait, we chose the dicotyledonous C_4 plant *Cleome gynandra*. For the biochemical characterization of C_4 in *C. gynandra*, we compared the enzyme activities of C_4 genes with those of the C_3 plant *Tarenaya hassleriana*. We detected an age-dependent plasticity of the composition of decarboxylating enzymes, where PEPCK activity increased from 52% of NADME activity in young leaves to 153% in old leaves. Furthermore, we showed the exclusively mitochondrial localization of the C_4 Aspartate Aminotransferase and propose a new biochemical model as an alternative to the textbook model of the NADME C_4 pathway (Manuscript 1).

To examine the development of the C_4 trait in *C. gynandra*, we conducted RNA sequencing analysis on five stages of developing leaves and other tissues in comparison to the C_3 plant *T. hassleriana*. We found a delayed expression of photosynthesis genes in the C_4 plant as compared to C_3 that is accompanied by a delay of Kranz differentiation. We concluded that the establishment of Kranz leaf anatomy at a later stage of development is causal for this delay. We also found that developmental modules of the root are transferred to Kranz-structured C_4 leaves, indicating that the leaf copies developmental cues from the concentric root tissue for bundle sheath and mesophyll differentiation. RNA sequencing in combination with nucleus size determination and flow cytometry helped us to find endoreduplication occurring in bundle sheath cells. We hypothesize that endoreduplication is essential to maintain high expression profiles of those metabolically active cells. Taken together, these results illustrate the development of the C_4 trait in leaves of *C. gynandra* (Manuscript 2).

Next, we analyzed the role of Golden2-like (Glk) transcription factors for the development of C_4 photosynthesis. Glk1 is the only expressed Glk in *C. gynandra* and shows no specificity for bundle sheath or mesophyll. We heterologously expressed *Cg*Glk1 in *A. thaliana* and demonstrated functional conservation between both plants by biochemical and phenotypic analysis. However, complementation was insufficient, since transcripts of PSII genes showed a variable, incomplete rescue in response to *Cg*Glk1 expression in *A. thaliana* with the weakest response being in oxygen

evolving complex genes. Future studies remain to determine, whether this effect is based on molecular evolution of *Cg*Glk1 (Manuscript-draft 3).

In the data from Manuscript 2, we found up-regulation of lipid transfer proteins (LTPs) in *C. gynandra* leaves. We functionally characterized the most abundant LTP, which is homologous to At1G12090 in *A. thaliana*, and found that it is a secretory protein that specifically binds cholesterol. High specificity for a single lipid species has not been demonstrated for a plant LTP before. Rare information about the role of sterols in C_4 photosynthesis do not allow for us to propose a hypothesis on the biological role of *Cg*LTP (Manuscript-draft 4).

To complement our findings on the development of C_4 photosynthesis in *C. gynandra*, we analyzed the transcriptome of the facultative C_4 sedge *Eleocharis retroflexa* in submerged (C_3) and terrestrial (C_4) growth. We found a high degree of phenotypic plasticity in secondarily submerged culms, which indicates that the C_4 trait is not completely shut off by a master regulator, but rather do C_4 genes follow more individual regulatory patterns (Manuscript-draft 5).

In summary, this thesis contributed to a better understanding of the development of C_4 photosynthesis and adaptation of regulatory proteins to the C_4 syndrome in the genus *Cleome*. It further points out the role of structural adaptation to C_4 photosynthesis and shows that C_4 genes can be regulated individually in response to the environment.

I.2 Zusammenfassung

C_4 Photosynthese ist ein Merkmal, das hohe Produktivität vermittelt und sie ist mehr als 66 Mal unabhängig voneinander entstanden. C_4 Pflanzen unterteilen die Photosynthese zwischen zwei verschiedenen Zelltypen. Die Biochemie des zentralen Stoffwechselzyklus, der diese Zellen miteinander verbindet, sowie die Blattanatomie sind hinreichend studiert. Im Gegensatz dazu existiert nur wenig Wissen über die zugrunde liegenden regulierenden Prozesse sowie die Entwicklung dieses Merkmals. Außerdem wurde die molekulare Natur der strukturellen Anpassung an die C_4 Photosynthese bisher nicht beschrieben. Das Ziel dieser Arbeit ist es, zu einem besseren Verständnis dieser globalen Anpassung einer Pflanze an die C_4 Photosynthese beizutragen.

Als Modellorganismus für die Analyse des C_4 Merkmals wählten wir die zweikeimblättrige C_4 Pflanze *Cleome gynandra* aus. Um die C_4 Photosynthese in *C. gynandra* biochemisch zu charakterisieren, verglichen wir die Enzymaktivitäten bekannter C_4 Gene mit denen der C_3 Pflanze *Tarenaya hassleriana*. Dabei entdeckten wir eine altersabhängige Umgestaltung in der Zusammensetzung der dekarboxylierenden Enzyme, bei der die PEPCK Aktivität von 52% der NADME Aktivität in jungen Blättern bis auf 153% in alten Blättern anstieg. Weiterhin orteten wir die in der C_4 Photosynthese genutzte Aspartat-Aminotransferase ausschließlich in Mitochondrien und schlagen ein von den Lehrbüchern abweichendes Modell für den NADME abhängigen C_4 Zyklus vor (Manuskript 1).

Um die Entwicklung des C_4 Merkmals in *C. gynandra* zu untersuchen, führten wir RNS Sequenzanalysen von fünf Entwicklungsstadien des Blattes und von anderen Geweben durch, und verglichen die Ergebnisse mit denen von der C_3 Pflanze *T. hassleriana*. Wir fanden eine verspätete Expression von Genen der Photosynthese in der C_4 Pflanze im Vergleich mit der C_3 Pflanze, die mit einer ebenfalls verspäteten Differenzierung der Kranz Anatomie übereinstimmte. Wir folgerten, dass die verspätete Etablierung der Kranz-Anatomie ursächlich für die verspätete Expression der Photosynthesegene ist. Kranz-Anatomie ist notwendig für die C_4 Photosynthese in den meisten höheren Pflanzen und daher können nicht-differenzierte Blätter, im Gegensatz zu C_3 Pflanzen, keine Photosynthese durchführen. Des Weiteren fanden wir heraus, dass Entwicklungsmodule aus der Wurzel in die kranzförmige Blattstruktur der C_4 Pflanze übertragen werden. Das weist darauf hin, dass das Blatt die Entwicklungssignale aus dem konzentrischen Wurzelgewebe kopiert und für die Entwicklung von Mesophyll- und Bündelscheidengewebe verwendet. Die RNS Sequenzanalyse half uns, in Verbindung mit der Größenbestimmung von Zellkernen und Durchflusszytometrie, Endoreduplikation in Bündelscheidenzellen nachzuweisen. Wir stellen die Hypothese auf, dass Endoreduplikation dazu beiträgt, die hohen Transkriptionsraten in diesen metabolisch sehr aktiven

Zellen zu ermöglichen. Zusammengenommen illustrieren diese Ergebnisse die Entwicklung des C_4 Merkmals in den Blättern von *C. gynandra* (Manuskript 2).

Anschließend analysierten wir die Rolle des Golden2-like (Glk) Transkriptionsfaktors in der Entwicklung der C_4 Photosynthese. Glk1 ist das einzige abgelesene Glk Gen in *C. gynandra* und ist nicht ausschließlich in Mesophyll- oder Bündelscheidengewebe vorhanden. Wir exprimierten *Cg*Glk1 heterolog in *A. thaliana* und zeigten Konservierung seiner Funktion zwischen den beiden Pflanzen durch biochemische und physiologische Analysen. Dennoch war die Komplementierung des Phänotyps unvollständig, weil die Transkriptmenge photosynthetischer Gene teils nur unvollständig wiederhergestellt wurde, wenn *Cg*Glk1 exprimiert wurde. Die schwächste Wiederherstellung sahen wir bei Genen des Sauerstoff bildenden Komplexes. Weitere Untersuchungen müssen noch zeigen, ob dieser Effekt auf molekularer Evolution des *Cg*Glk1 Gens beruht (Manuskript-Entwurf 3).

In den Daten von Manuskript 2 fanden wir Lipid Transfer Proteine (LTPs) in Blättern von *C. gynandra* hochreguliert. Wir charakterisierten die Funktion des häufigsten LTP, welches homolog zu At1G12090 in *A. thaliana* ist. Dabei fanden wir heraus, dass es ein sekretorisches Protein ist, welches spezifisch an Cholesterin bindet. Eine hohe Spezifität für lediglich ein einziges Lipid wurde bisher noch nicht für ein LTP einer Pflanze gezeigt. Eine biologische Funktion für *Cg*LTP können wir jedoch aufgrund des Mangels an Informationen über die Rolle von Sterolen in der C_4 Photosynthese nicht vorschlagen (Manuskript-Entwurf 4).

Um unsere Ergebnisse, die die Entwicklung der C_4 Photosynthese in *C. gynandra* betreffen, zu komplementieren, analysierten wir das Transkriptom der fakultativen C_4 Segge *Eleocharis retroflexa* während des Wachstums unter Wasser (C_3) und an Land (C_4). Wir fanden ein hohes Maß an Plastizität des C_4 Phänotyps bei sekundär untergetauchten Halmen, was darauf hinweist, dass das C_4 Merkmal nicht vollständig durch einen Hauptschalter abgestellt wird, sondern dass einzelne Gene individueller Regulation unterliegen (Manuskript-Entwurf 5).

Zusammengefasst hat diese Arbeit zu einem besseren Verständnis der Entwicklung der C_4 Photosynthese und der Anpassung von regulatorischen Proteinen an das C_4 Syndrom in der Gattung *Cleome* geführt. Des Weiteren verdeutlicht sie die Rolle der strukturellen Anpassung an die C_4 Photosynthese und zeigt, dass C_4 Gene in Anpassung an ihre Umwelt individuell reguliert werden können.

II. Introduction

Life on earth is almost exclusively powered by solar energy that is used by autotrophic organisms to reduce carbon dioxide (CO_2) and assimilate energy-rich sugars via oxygenic photosynthesis. Oxygenic photosynthesis arose in microbes about 2.4 billion years ago (Farquhar et al., 2011) and the core reaction of photosynthesis is highly conserved in all plants. During oxygenic photosynthesis, one molecule of CO_2 and one molecule of water (H_2O) are required to fix the CO_2's carbon atom into a carbohydrate and as a byproduct one molecule of O_2 gets produced. However, to increase the efficiency of photosynthesis, plants evolved the intricate photosynthetic apparatus that optimizes the reaction conditions of the core reaction. One function of the photosynthetic apparatus is to concentrate light energy at the site where H_2O is deprotonated and the electrons for CO_2 fixation are liberated. This is achieved by the photosynthetic antenna, a protein complex that is located in the thylakoid membrane of the chloroplast and hosts numerous pigments (Liu et al., 2004). When these pigments are encountered by photons, electrons are excited to a higher energy level and subsequently pass this energy to neighboring pigments until they are collected at a central chlorophyll a pigment to drive an electron transport chain (Kühlbrandt et al., 1994). A second important function of the photosynthetic apparatus is the controlled breakdown of H_2O molecules into protons, electrons and molecular oxygen and the electron transport itself so that electrons can be loaded to $NADP^+$ and serve as a mobile energy equivalent (Haehnel, 1984). The complexity of photosynthetic carbon fixation in plants illustrates the importance of these reactions, but regardless of very detailed knowledge about the function of the photosynthetic apparatus, only a single family of transcription factors regulating its assembly has been found, namely Golden2-like 1 and 2 (Glk1 and Glk2) (Hall et al., 1998; Waters et al., 2009). Knockout of both Glk genes leads to a severely impaired photosynthetic apparatus in *Arabidopsis thaliana*.

The energy equivalents that are produced at the photosynthetic apparatus get utilized in the CO_2 fixation reaction, which does not occur at the photosynthetic apparatus but in the chloroplast stroma (Peltier et al., 2006). This causes spatial separation of CO_2 fixation from the O_2 producing H_2O breakdown. Separation of these reactions is particularly important for efficient CO_2 fixation, since the enzyme that carboxylates the acceptor molecule Ribulose-1,5-bisphosphate (RuBP) can also accept O_2 as a substrate. The enzyme is therefore called Ribulose-1,5-bispohsphate carboxylase oxygenase (RubisCO) and by means of its crucial function the most abundant enzyme on earth (Ellis, 1979; Raven, 2013). When RubisCO first evolved, the global atmosphere mainly consisted of CO_2, which allowed for a low substrate specificity of the enzyme but the rise of autotrophic organisms led to massive photosynthetic carbon fixation and the partial pressure of O_2 rose (Kasting, 1987). By this transformation of the atmospheric composition, autotrophs challenged their own carbon fixing enzyme

to minimize the oxygenation reaction, but despite slow catalysis and low substrate specificity, RubisCO is probably already perfectly optimized under ambient air conditions (Tcherkez et al., 2006). The competition of O_2 and CO_2 for the active site of RubisCO causes an oxygenation in one out of five RubisCO reactions at recent atmospheric composition in the model plant *Glycine max* (21% O_2 and 0.04% CO_2) (Laing et al., 1974). The oxygenation reaction of RubisCO yields 2-phosphoglycolate (2-PG), which is phytotoxic and has to be recycled for a loss of energy via the photorespiratory pathway (Ogren, 1984). The energetic demand of photorespiration became the driving force for the evolution of carbon concentrating mechanisms (CCMs) at the site of RubisCO to increase the enzyme's efficiency in a largely oxidizing environment (Raven et al., 2008).

Evolution of Carbon Concentrating Mechanisms

CCMs accumulate CO_2 at the site of RubisCO to saturate its carboxylation reaction and they can be found in cyanobacteria, algae and in numerous embryophytes (Raven et al., 2008). CCMs are complex traits and affect metabolism as well as cellular and subcellular architecture. It was shown, however, that these traits evolved stepwise and that each incremental step described a selective advantage for the organism (Heckmann et al., 2013). Carbon concentration at the site of RubisCO can occur in either a temporal or a spatial manner and it can be achieved by both avoidance of CO_2 loss and active CO_2 enrichment. The enrichment of CO_2 is usually accompanied by reduction of oxygen evolving reactions when or where RubisCO is active (Raven et al., 2008). The best studied CCM is called C_4 photosynthesis, as the first product of carbon fixation is a four-carbon compound instead of 3-phosphoglycerate.

The Central Pathway of C_4 Photosynthesis

C_4 photosynthesis was discovered in 1966 by MD Hatch and CR Slack (Hatch and Slack, 1966). C_4 plants keep the majority of their RubisCO in a central leaf tissue, the bundle sheath. Oxygen evolving reactions, in contrast, are located in peripheral mesophyll cells (Edwards et al., 2001; Leegood, 2002; Majeran et al., 2005; Majeran et al., 2008). Carbon enrichment occurs via a metabolic shuttle from mesophyll to bundle sheath cells. Therefore, CO_2 is fixed by phosphoenolpyruvate-carboxylase (PEPC) in mesophyll, yielding the C_4 acid oxaloacetate. In contrast to RubisCO, PEPC is insensitive to O_2 and is thus not inhibited by ambient oxygen levels (Bowes and Ogren, 1972). After CO_2 fixation, oxaloacetate or a similar C_4-compound is transported into the bundle sheath tissue following a concentration gradient and probably also in an active manner (Weiner et al., 1988; Sowiński et al., 2008). In the bundle sheath, CO_2 is released from the C_4-compound by a decarboxylating enzyme. Three decarboxylating enzymes have been identified in C_4 photosynthesis: NAD-dependent malic enzyme (NADME), NADP-dependent malic enzyme (NADPME) and phosphoenolpyruvate-

carboxykinase (PEPCK). C_4 plants can be classified by the mainly used decarboxylating enzyme (Hatch, 1988). However, it was shown that multiple decarboxylating enzymes can contribute to C_4 photosynthesis in a single species and that recruitment of different enzymes occurs in an age-related manner (Wingler et al., 1999; Furbank, 2011; Sommer et al., 2012; Muhaidat and McKown, 2013; Külahoglu et al., 2014; Wang et al., 2014). After CO_2 is released by a decarboxylating enzyme, the remaining C_3 compound is regenerated and transported back to the mesophyll cell for a new cycle of fixation, release and regeneration. The direct and stoichiometric role in photosynthetic carbon fixation makes the C_4 cycle essential for C_4 plants and its intermediates accumulate to concentrations as high as 18 mM (Hatch and Osmond, 1976). By this CCM, the oxygenation reaction of RubisCO is strongly reduced (Gowik and Westhoff, 2011). This avoids photorespiration and leads to numerous environmental advantages of C_4 photosynthesis over C_3 photosynthesis, like greater crop yields and increased use efficiencies of water and nitrogen (Schmitt and Edwards, 1981; Sage, 2004; Zhu et al., 2010; Covshoff and Hibberd, 2012). These advantages are most distinct at high temperatures, since the specificity of RubisCO for CO_2 decreases with increasing temperature (Bernacchi et al., 2001; Zhu et al., 2008). However, establishment of the C_4 cycle was also a costly process. Two extra ATPs are required for CO_2 pre-fixation in C_4 plants (Zhu et al., 2008) and anatomical adaptations of the leaves are necessary for efficient C_4 photosynthesis. Commonly, C_4 plants display an increased vein density, so that two veins are separated by only two bundle sheath cells and two mesophyll cells. Cross sections show that the mesophyll cells surround the bundle sheath tissue in a concentric shape (Hattersley, 1984; Muhaidat et al., 2007). Due to this concentric organization around veins, the leaf structure of C_4 plants is referred to as Kranz-anatomy.

Enzymes and metabolites of the central C_4 cycle were excessively studied in the last century (Hatch, 1988). The biochemistry of anatomical adaptations and regulators of C_4 photosynthesis remain largely unknown, however.

Beyond Central C_4 Metabolism – Leaf Anatomy and Regulators of Photosynthesis

The organization of Kranz anatomy requires precise coordination of cell identities. This is demonstrated by predominant conformity of the existence of a single mesophyll cell layer around each bundle sheath cell layer in C_4 plants (Hattersley, 1984; Muhaidat et al., 2007). Only recently, however, first insights into the regulation of Kranz development emerged. In maize, the root transcription factor SCARECROW is involved in leaf vein formation, as defective mutants display fewer veins than the wildtype (Slewinski et al., 2012). It was subsequently proposed that bundle sheath architecture is an extension of developmental cues of root endodermis tissue to the leaf (Slewinski, 2013). A comparative transcriptome atlas of the C_4 plant *Cleome gynandra* and the C_3 plant *Tarenaya hassleriana* affirmed this hypothesis. It revealed a cluster of genes that show expression in roots of the

C_3 plant in leaves of the C_4 plant. In *Arabidopsis thaliana*, genes of this cluster were mostly expressed in the root endodermis or cortex tissue, which are homologs of bundle sheath and mesophyll tissue in leaves, respectively (Külahoglu et al., 2014). Since most of these genes are not functionally characterized to date, their assignment to roles in C_4 photosynthesis was not yet possible. Besides an increased expression of the root gene cluster, a number of lipid transfer proteins (LTPs) were strongly up-regulated in leaf tissue of the C_4 plant. LTPs are small proteins with a hydrophobic pocket that can bind and transport lipids *in vitro* (Kader, 1996). The family of LTPs contains 49 members in *A. thaliana* and only few of them were functionally characterized (Kader, 1996; Boutrot et al., 2008). LTPs commonly contain a secretory target peptide and are thus located to the extracellular space (Mundy and Rogers, 1986; Sterk et al., 1991; Kader, 1996). Contribution to the formation of extracellular matrices was hypothesized as a role of extracellular LTPs (Sterk et al., 1991; Douliez et al., 2000). However, a role of LTPs in C_4 photosynthesis was not yet proposed. The regulation of vein density, bundle sheath development and tightly controlled mesophyll cell formation in C_4 plants thus remains unknown.

To understand Kranz anatomy from a functional point of view, phenotypic analyses were performed. Cross sections of C_4 leaves showed that some monocotyledonous C_4 species exhibit a suberized cell wall lamella around bundle sheath cells (Eastman et al., 1988; Nelson, 2011). It was hypothesized that this suberin layer causes a decreased permeability for CO_2 and can thus minimize diffusion from the CO_2-enriched bundle sheath to other tissues (Furbank et al., 1989; Jenkins et al., 1989). A comparable structure is lacking in dicotyledonous C_4 plants. In dicotyledonous NADME plants, CO_2 diffusion is minimized by a centripetal arrangement of mitochondria and chloroplasts (von Caemmerer and Furbank, 2003). Mitochondria and chloroplasts are the site of CO_2 liberation and fixation, respectively and by their centripetal arrangement, the diffusion distance to the mesophyll tissue is maximized (von Caemmerer and Furbank, 2003; Marshall et al., 2007). The protection of the bundle sheath from CO_2 diffusion losses poses an anatomical challenge to C_4 plants as it is in contrast to the high demand for exchange of C_4 cycle intermediates between the tissues. C_4 plants accomplish the metabolite exchange via a high number of plasmodesmata at the bundle sheath-mesophyll interface (Botha, 1992). Plasmodesmata are cytosolic connections between two cells, which allow for passive movement of small molecules (water, sugars, small organic acids). Feedback regulation of plasmodesmata aperture by the redox state of organelles was demonstrated, while other regulatory mechanisms (light, hormones) were detected but not yet characterized (Burch-Smith and Zambryski, 2012). The regulatory process that orchestrates plasmodesmata predominantly to the bundle sheath-mesophyll interface remains to be elucidated, too.

Besides anatomical adaptations to Kranz anatomy, cell specific regulation of the photosynthetic apparatus poses a key challenge in the development of C_4 architecture. The reduction

of oxygen evolution is facilitated by a decrease in photosystem II core genes in BS cells (Majeran et al., 2008). This phenotype is accompanied by an increase in cyclic electron transport due to a lack of electrons provided by water splitting reactions (Majeran and van Wijk, 2009; Bräutigam et al., 2011). In maize, the *bundle sheath defective* (*bsd*) 1 mutant shows a chlorotic phenotype that manifests most prominently in bundle sheath cells (Hall et al., 1998). The mutant was allelic for the Golden2 transcription factor and Golden2 mRNA was found to localize primarily to the bundle sheath (Rossini et al., 2001). A second Golden2 gene, namely Golden2-like 1 (Glk1) was found to be expressed primarily in mesophyll cells. The C_3 plant *A. thaliana* contains two Glk genes as well, but their expression pattern was shown to overlap (Rossini et al., 2001). Single *glk* mutants show no visible phenotype, while the *glk1 glk2* double mutant appears uniformly pale green (Fitter et al., 2002). It was thus concluded that Glk function is redundant in C_3 plants. A transcriptome study of *glk* mutants in *A. thaliana* is in line with this conclusion and furthermore revealed the targets of GLK proteins in a C_3 plant. The main targets were nuclear encoded members of the photosynthetic apparatus and genes of chlorophyll biosynthesis (Waters et al., 2009). The global effect of the *glk*1 *glk*2 knockout on the *A. thaliana* transcriptome remains to be determined. Due to the compartmentalization of Glk genes in the C_4 plant maize, it was concluded that the individual Glk genes specialized to cell-type specific differentiation processes (Rossini et al., 2001). Glk expression in the dicotyledonous C_4 plant *C. gynandra*, however, is different from the cell-type specific expression in maize (Wang et al., 2013). *C. gnynadra* expresses only Glk1, while Glk2 is almost completely silenced. *Cg*Glk1 is predominantly expressed in mesophyll cells, but a notable amount of transcript was also detected in bundle sheath cells (Wang et al., 2013). Since *C. gynandra* bundle sheath chloroplasts are, unlike in maize, not enlarged and agranal (Marshall et al., 2007), it was hypothesized that Glk genes mediate chloroplast dimorphism in maize (Wang et al., 2013). However, while cell-specific Glk expression is lacking in *C. gynandra*, it performs C_4 photosynthesis and transcript levels of cyclic electron transport genes are enhanced (Bräutigam et al., 2011; Sommer et al., 2012). It was not shown to date, which role *Cg*Glk1 plays for the establishment of the C_4 syndrome in *C. gynandra*. Other transcription factors regulating photosynthetic gene expression are yet to be determined.

Cleome as a Model for the Evolution of C_4 Photosynthesis

The genus *Cleome* belongs to the family of *Cleomaceae* and contains dicotyledonous plants. *Cleome* is one of over 60 evolutionary origins of C_4 photosynthesis, and as a consequence it contains C_3 species, as well as C_3-C_4 intermediates and C_4 species (Marshall et al., 2007). The family *Cleomaceae* is a sister clade of the family of Brassicaceae, which *A. thaliana* is a member of (Schranz and Mitchell-Olds, 2006). This makes the C_4 origin in the *Cleomaceae* the closest C_4 origin to the model species of plant genetics. Short evolutionary distance minimizes random or speciation-related

effects of molecular evolution, which simplifies heterologous expression and functional annotation of genes. Thus, *Cleome* provides a favorable model for molecular studies on C_4 photosynthesis (Brown et al., 2005). For comparative analyses between a C_3 and a C_4 species, *Tarenaya hassleriana* (also known as *Cleome spinosa*) and *Gynandropsis gynandra* (also known as *Cleome gynandra*) were chosen, respectively. Transcriptomes of mature leaves and a transcriptome atlas including leaf development of both species were published earlier (Bräutigam et al., 2011; Külahoglu et al., 2014). In anatomical and biochemical studies, *C. gynandra* was confirmed an NADME-type C_4 species with side activity of PEPCK (Marshall et al., 2007; Sommer et al., 2012). For *T. hassleriana*, the full genome was published recently, which further simplified the classification of large transcriptome datasets (Cheng et al., 2013). *C. gynandra* thus is a favorable species for further investigation of the C_4 syndrome.

Complete avoidance of speciation effects by analyzing C_3 and C_4 in a single species

In some plant species, C_4 photosynthesis is a facultative trait. Among these species are sedges of the genus *Eleocharis* (Ueno et al., 1988). When submerged in water, these plants perform C_4 photosynthesis and when grown terrestrial they switch to C_4 photosynthesis. This existence of two photosynthetic types in a single organism allows for comparative analysis without speciation-dependent noise in the results. However, the change from an aquatic to a terrestrial environment can cause secondary effects that are unrelated to the mode of photosynthesis. Furthermore, no genome of an *Eleocharis* species has been sequenced yet and thus, the next available source of sequence information is the genome of *Setaria italica*. Consequently, despite some disadvantages, studying C_4 photosynthesis is well suited as a complementary approach to other comparative analyses of C_3 and C_4 species.

References

Bernacchi, C.J., Singsaas, E.L., Pimentel, C., Portis Jr, A.R., and Long, S.P. (2001). Improved temperature response functions for models of Rubisco-limited photosynthesis. Plant, Cell & Environment **24,** 253-259.

Botha, C.E.J. (1992). Plasmodesmatal distribution, structure and frequency in relation to assimilation in C3 and C4 grasses in southern Africa. Planta **187,** 348-358.

Boutrot, F., Chantret, N., and Gautier, M.-F. (2008). Genome-wide analysis of the rice and arabidopsis non-specific lipid transfer protein (nsLtp) gene families and identification of wheat nsLtp genes by EST data mining. BMC Genomics **9,** 86.

Bowes, G., and Ogren, W.L. (1972). Oxygen Inhibition and Other Properties of Soybean Ribulose 1,5-Diphosphate Carboxylase. Journal of Biological Chemistry **247,** 2171-2176.

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

Brown, N.J., Parsley, K., and Hibberd, J.M. (2005). The future of C4 research – maize, Flaveria or Cleome? Trends Plant Sci. **10,** 215-221.

Burch-Smith, T.M., and Zambryski, P.C. (2012). Plasmodesmata Paradigm Shift: Regulation from Without Versus Within. Annual Review of Plant Biology **63,** 239-260.

Cheng, S., van den Bergh, E., Zeng, P., Zhong, X., Xu, J., Liu, X., Hofberger, J., de Bruijn, S., Bhide, A.S., Kuelahoglu, C., Bian, C., Chen, J., Fan, G., Kaufmann, K., Hall, J.C., Becker, A., Brautigam, A., Weber, A.P., Shi, C., Zheng, Z., Li, W., Lv, M., Tao, Y., Wang, J., Zou, H., Quan, Z., Hibberd, J.M., Zhang, G., Zhu, X.G., Xu, X., and Schranz, M.E. (2013). The Tarenaya hassleriana genome provides insight into reproductive trait and genome evolution of crucifers. The Plant cell **25,** 2813-2830.

Covshoff, S., and Hibberd, J.M. (2012). Integrating C4 photosynthesis into C3 crops to increase yield potential. Current Opinion in Biotechnology **23,** 209-214.

Douliez, J.P., Michon, T., Elmorjani, K., and Marion, D. (2000). Mini Review: Structure, Biological and Technological Functions of Lipid Transfer Proteins and Indolines, the Major Lipid Binding Proteins from Cereal Kernels. Journal of Cereal Science **32,** 1-20.

Eastman, P.A., Dengler, N., and Peterson, C. (1988). Suberized bundle sheaths in grasses (Poaceae) of different photosynthetic types I. Anatomy, ultrastructure and histochemistry. Protoplasma **142,** 92-111.

Edwards, G.E., Franceschi, V.R., Ku, M.S.B., Voznesenskaya, E.V., Pyankov, V.I., and Andreo, C.S. (2001). Compartmentation of photosynthesis in cells and tissues of C4 plants. Journal of Experimental Botany **52,** 577-590.

Ellis, R.J. (1979). The most abundant protein in the world. Trends in Biochemical Sciences **4,** 241-244.

Farquhar, J., Zerkle, A., and Bekker, A. (2011). Geological constraints on the origin of oxygenic photosynthesis. Photosynthesis Research **107,** 11-36.

Fitter, D.W., Martin, D.J., Copley, M.J., Scotland, R.W., and Langdale, J.A. (2002). GLK gene pairs regulate chloroplast development in diverse plant species. The Plant journal : for cell and molecular biology **31,** 713-727.

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

Furbank, R.T., Jenkins, C.L., and Hatch, M.D. (1989). CO2 concentrating mechanism of C4 photosynthesis permeability of isolated bundle sheath cells to inorganic carbon. Plant Physiol. **91,** 1364-1371.

Gowik, U., and Westhoff, P. (2011). The Path from C3 to C4 Photosynthesis. Plant Physiol. **155,** 56-63.

Haehnel, W. (1984). Photosynthetic electron transport in higher plants. Annual Review of Plant Physiology **35,** 659-693.

Hall, L.N., Rossini, L., Cribb, L., and Langdale, J.A. (1998). GOLDEN 2: A Novel Transcriptional Regulator of Cellular Differentiation in the Maize Leaf. The Plant Cell Online **10,** 925-936.

Hatch, M. (1988). C4 photosynthesis, a unique blend of modified biochemistry, anatomy, and ultrastructure. Biochim. Biophys. Acta **895,** 81-106.

Hatch, M., and Osmond, C. (1976). Compartmentation and transport in C4 photosynthesis. In Transport in Plants III (Springer), pp. 144-184.

Hatch, M.D., and Slack, C.R. (1966). Photosynthesis by sugar-cane leaves. Biochem. J **101,** 103-111.

Hattersley, P. (1984). Characterization of C4 type leaf anatomy in grasses (Poaceae). Mesophyll: bundle sheath area ratios. Annals of Botany **53,** 163-180.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, Andreas P.M., and Lercher, Martin J. (2013). Predicting C4 Photosynthesis Evolution: Modular, Individually Adaptive Steps on a Mount Fuji Fitness Landscape. Cell **153,** 1579-1588.

Jenkins, C.L.D., Furbank, R.T., and Hatch, M.D. (1989). Inorganic Carbon Diffusion between C(4) Mesophyll and Bundle Sheath Cells: Direct Bundle Sheath CO(2) Assimilation in Intact Leaves in the Presence of an Inhibitor of the C(4) Pathway. Plant Physiol. **91,** 1356-1363.

Kader, J.-C. (1996). LIPID-TRANSFER PROTEINS IN PLANTS. Annu. Rev. Plant Physiol. Plant Molec. Biol. **47,** 627-654.

Kasting, J.F. (1987). Theoretical constraints on oxygen and carbon dioxide concentrations in the Precambrian atmosphere. Precambrian Research **34,** 205-229.

Kühlbrandt, W., Wang, D., and Fujiyoshi, Y. (1994). Atomic model of plant light-harvesting complex by electron crystallography. Nature **367,** 614-621.

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

Laing, W.A., Ogren, W.L., and Hageman, R.H. (1974). Regulation of Soybean Net Photosynthetic CO2 Fixation by the Interaction of CO2, O2, and Ribulose 1,5-Diphosphate Carboxylase. Plant Physiol. **54,** 678-685.

Leegood, R.C. (2002). C4 photosynthesis: principles of CO2 concentration and prospects for its introduction into C3 plants. Journal of Experimental Botany **53,** 581-590.

Liu, Z., Yan, H., Wang, K., Kuang, T., Zhang, J., Gui, L., An, X., and Chang, W. (2004). Crystal structure of spinach major light-harvesting complex at 2.72[thinsp]A resolution. Nature **428,** 287-292.

Majeran, W., and van Wijk, K.J. (2009). Cell-type-specific differentiation of chloroplasts in C4 plants. Trends Plant Sci. **14,** 100-109.

Majeran, W., Cai, Y., Sun, Q., and van Wijk, K.J. (2005). Functional Differentiation of Bundle Sheath and Mesophyll Maize Chloroplasts Determined by Comparative Proteomics. The Plant Cell Online **17,** 3111-3140.

Majeran, W., Zybailov, B., Ytterberg, A.J., Dunsmore, J., Sun, Q., and van Wijk, K.J. (2008). Consequences of C4 Differentiation for Chloroplast Membrane Proteomes in Maize Mesophyll and Bundle Sheath Cells. Molecular & Cellular Proteomics **7,** 1609-1638.

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

Muhaidat, R., and McKown, A.D. (2013). Significant involvement of PEP-CK in carbon assimilation of C4 eudicots. Annals of Botany **111,** 577-589.

Muhaidat, R., Sage, R.F., and Dengler, N.G. (2007). Diversity of Kranz anatomy and biochemistry in C4 eudicots. Am. J. Bot. **94,** 362-381.

Mundy, J., and Rogers, J.C. (1986). Selective expression of a probable amylase/protease inhibitor in barley aleurone cells: Comparison to the barley amylase/subtilisin inhibitor. Planta **169,** 51-63.

Nelson, T. (2011). The grass leaf developmental gradient as a platform for a systems understanding of the anatomical specialization of C4 leaves. Journal of Experimental Botany **62,** 3039-3048.

Ogren, W.L. (1984). Photorespiration: pathways, regulation, and modification. Annual Review of Plant Physiology **35,** 415-442.

Peltier, J.-B., Cai, Y., Sun, Q., Zabrouskov, V., Giacomelli, L., Rudella, A., Ytterberg, A.J., Rutschow, H., and van Wijk, K.J. (2006). The Oligomeric Stromal Proteome of Arabidopsis thaliana Chloroplasts. Molecular & Cellular Proteomics **5,** 114-133.

Raven, J.A. (2013). Rubisco: still the most abundant protein of Earth? New Phytol. **198,** 1-3.

Raven, J.A., Cockell, C.S., and De La Rocha, C.L. (2008). The evolution of inorganic carbon concentrating mechanisms in photosynthesis.

Rossini, L., Cribb, L., Martin, D.J., and Langdale, J.A. (2001). The Maize Golden2 Gene Defines a Novel Class of Transcriptional Regulators in Plants. The Plant Cell Online **13,** 1231-1244.

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

Schmitt, M.R., and Edwards, G.E. (1981). Photosynthetic Capacity and Nitrogen Use Efficiency of Maize, Wheat, and Rice: A Comparison Between C3 and C4 Photosynthesis. Journal of Experimental Botany **32,** 459-466.

Schranz, M.E., and Mitchell-Olds, T. (2006). Independent Ancient Polyploidy Events in the Sister Families Brassicaceae and Cleomaceae. The Plant Cell Online **18,** 1152-1165.

Slewinski, T.L. (2013). Using evolution as a guide to engineer Kranz-type C4 photosynthesis. Frontiers in plant science **4**.

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

Sommer, M., Bräutigam, A., and Weber, A. (2012). The dicotyledonous NAD malic enzyme C4 plant Cleome gynandra displays age-dependent plasticity of C4 decarboxylation biochemistry. Plant Biology **14,** 621-629.

Sowiński, P., Szczepanik, J., and Minchin, P.E.H. (2008). On the mechanism of C4 photosynthesis intermediate exchange between Kranz mesophyll and bundle sheath cells in grasses. Journal of Experimental Botany **59,** 1137-1147.

Sterk, P., Booij, H., Schellekens, G.A., Van Kammen, A., and De Vries, S.C. (1991). Cell-specific expression of the carrot EP2 lipid transfer protein gene. The Plant Cell Online **3,** 907-921.

Tcherkez, G.G.B., Farquhar, G.D., and Andrews, T.J. (2006). Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. Proceedings of the National Academy of Sciences **103,** 7246-7251.

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.

von Caemmerer, S., and Furbank, R. (2003). The C4 pathway: an efficient CO2 pump. Photosynthesis Research **77,** 191-207.

Wang, P., Fouracre, J., Kelly, S., Karki, S., Gowik, U., Aubry, S., Shaw, M.K., Westhoff, P., Slamet-Loedin, I.H., Quick, W.P., Hibberd, J.M., and Langdale, J.A. (2013). Evolution of GOLDEN2-LIKE gene function in C(3) and C(4) plants. Planta **237,** 481-495.

Wang, Y., Bräutigam, A., Weber, A.P.M., and Zhu, X.-G. (2014). Three distinct biochemical subtypes of C4 photosynthesis? A modelling analysis. Journal of Experimental Botany.

Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J., and Langdale, J.A. (2009). GLK Transcription Factors Coordinate Expression of the Photosynthetic Apparatus in Arabidopsis. The Plant Cell Online **21,** 1109-1128.

Weiner, H., Burnell, J.N., Woodrow, I.E., Heldt, H.W., and Hatch, M.D. (1988). Metabolite Diffusion into Bundle Sheath Cells from C4 Plants: Relation to C4 Photosynthesis and Plasmodesmatal Function. Plant Physiol. **88,** 815-822.

Wingler, A., Walker, R.P., Chen, Z.-H., and Leegood, R.C. (1999). Phosphoenolpyruvate Carboxykinase Is Involved in the Decarboxylation of Aspartate in the Bundle Sheath of Maize. Plant Physiol. **120,** 539-546.

Zhu, X.-G., Long, S.P., and Ort, D.R. (2008). What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? Current Opinion in Biotechnology **19,** 153-159.

Zhu, X.-G., Long, S.P., and Ort, D.R. (2010). Improving photosynthetic efficiency for greater yield. Annual Review of Plant Biology **61,** 235-261.

III. Manuscripts

Manuscript 1

The dicotyledonous NAD malic enzyme C_4 plant *Cleome gynandra* displays age-dependent plasticity of C_4 decarboxylation biochemistry

plant biology

Plant Biology ISSN 1435-8603

RESEARCH PAPER

The dicotyledonous NAD malic enzyme C_4 plant Cleome gynandra displays age-dependent plasticity of C_4 decarboxylation biochemistry

M. Sommer, A. Bräutigam & A. P. M. Weber

Institute of Plant Biochemistry, Heinrich-Heine-University, Düsseldorf, Germany

Keywords

Aspartate aminotransferase; C_4 photosynthesis; *Cleome*; NAD-dependent malic enzyme.

Correspondence

A. Weber, Institute of Plant Biochemistry, Heinrich-Heine-Universität, Gebäude 26.03, Ebene 01, Universitätsstraße 1, 40225 Düsseldorf, Germany.

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

Editor

R. Leegood

Received: 22 August 2011; Accepted: 19 October 2011

doi:10.1111/j.1438-8677.2011.00539.x

ABSTRACT

The C_4 photosynthetic pathway enriches carbon dioxide in the vicinity of Rubisco, thereby enabling plants to assimilate carbon more efficiently. Three canonical subtypes of C_4 exist, named after their main decarboxylating enzymes: NAD-dependent malic enzyme type, NADP-dependent malic enzyme type and phospho*enol*pyruvate carboxykinase type. *Cleome gynandra* is known to perform NAD-ME type C_4 photosynthesis. To further assess the mode of C_4 in *C. gynandra* and its manifestation in leaves of different age, total enzyme activities of eight C_4-related enzymes and the relative abundance of 31 metabolites were measured. *C. spinosa* was used as a C_3 control. *C. gynandra* was confirmed as an NAD-ME type C_4 plant in mid-aged leaves, whereas a mixed NAD-ME and PEPCK type was observed in older leaves. Young leaves showed a C_3-C_4 intermediate state with respect to enzyme activities and metabolite abundances. Comparative transcriptome analysis of mid-aged leaves of *C. gynandra* and *C. spinosa* showed that the transcript of only one aspartate aminotransferase (AspAT) isoform is highly abundant in *C. gynandra*. However, the canonical model of the NAD-ME pathway requires two AspATs, a mitochondrial and a cytosolic isoform. Surprisingly, our results indicate the existence of only one highly abundant AspAT isoform. Using GFP-fusion, this isozyme was localised exclusively to mitochondria. We propose a revised model of NAD-ME type C_4 photosynthesis in *C. gynandra*, in which both AspAT catalysed reactions take place in mitochondria and PEPCK catalyses an alternative decarboxylating pathway.

INTRODUCTION

Green plants produce organic matter from gaseous carbon dioxide (CO_2) using photosynthesis. While the core reactions of the Calvin-Benson cycle are conserved, different carbon enrichment mechanisms have evolved to improve the efficiency of carbon fixation. Although many seed plants assimilate CO_2 using only the Calvin-Benson cycle (*i.e.*, the C_3 pathway), employing ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) as the sole CO_2 assimilating enzyme, some plants use C_4 photosynthesis (Hatch & Slack 1966). C_4 photosynthesis has evolved independently multiple times from C_3 photosynthesis and leads to accumulation of CO_2 in the bundle sheath layer, where regular Rubisco fixation takes place. This makes C_4 plants more efficient at carbon assimilation, leading to higher water use efficiency (Black 1973), higher nitrogen use efficiency and/or faster accumulation of biomass. C_4 photosynthesis is a remarkable example of convergent evolution. The pathway has evolved independently over 60 times and independent origins exist both in eudicots and monocots (Sage 2004; Sage *et al.* 2011).

In principle, C_4 photosynthesis needs a primary fixation enzyme, which is specific for carbon assimilation and does not react with O_2, as is the case for Rubisco. In all known C_4 species, phospho*enol*pyruvate (PEP) carboxylase (PEPC) is the primary CO_2-fixing enzyme. The CO_2 is transported to the site of the Calvin cycle in the form of a C_4 acid and released from the C_4 acid by a decarboxylation enzyme, either PEP carboxykinase (PEPCK), NAD-dependent malic enzyme (NAD-ME) or NADP-dependent malic enzyme (NADP-ME) (Hatch 1987). Traditionally, C_4 plants are classified according to which is the major decarboxylation enzyme. Recently, however, it was hypothesised that plasticity exists in the decarboxylation biochemistry in response to environmental cues (Furbank 2011). The dicot *Cleome gynandra* is classified as a NAD-ME plant (Marshall *et al.* 2007). The NAD-ME pathway was primarily elaborated using the monocot model *Panicum miliaceum* and it was assumed to also operate in dicotyledonous NAD-ME C_4 plants, such as *C. gynandra* (Bräutigam *et al.* 2011a). Based on the *P. miliaceum* model, CO_2 is fixed into oxaloacetate (OAA), which is subsequently converted to Asp by cytosolic aspartate aminotransferase (AspAT) in mesophyll cells. Asp is then transferred to the bundle sheath, where it is transaminated to OAA, reduced to malate, and finally decarboxylated, releasing the CO_2 and pyruvate. Pyruvate is transaminated to alanine, thus serving as the terminal amino group acceptor for the AspAT reaction. The C_3 amino acid returns to the mesophyll, where it is converted over several steps into PEP, thus providing the precursor for a new round of carboxylation and decarboxylation. The transfer of a C_4 amino acid to the bundle sheath and the return of a C_3 amino acid to the mesophyll cells

balances the amino group transfer. The canonical model of NAD-ME type C_4 photosynthesis assumes that two isoforms of AspAT exist: one AspAT in the mesophyll cell cytosol and one AspAT in the bundle sheath mitochondria (as summarised in Hatch 1987). In the NAD-ME plant *P. miliaceum*, total AspAT activity is higher than in C_3 plants and distributed equally between mesophyll and bundle sheath (Hatch & Mau 1973). The AspAT activity in the bundle sheath is localised to mitochondria, while the localisation of mesophyll AspAT is not organelle associated (Hatch & Mau 1973). In a recent comparative transcriptome analysis of *C. spinosa* and *C. gynandra* using mRNA-Seq, only one out of five AspAT isozymes, a mitochondrial AspAT, was found to be upregulated in C_4 plants (Bräutigam *et al.* 2011a).

Compared to C_4 monocot model species, relatively little is known about dicotyledonous C_4 models, especially the new model *C. gynandra* (Brown *et al.* 2005). For example, in *Z. mays* (maize) it has been shown that C_4 photosynthesis develops with leaf age and with light exposure (Bassi & Passera 1982). Cross-sections of leaves show that Kranz anatomy develops during maturation, and it was concluded that C_4 metabolism develops during maturation of leaves and continuously takes over primary carbon fixation (Miranda *et al.* 1981). In the dicot *Flaveria trinervia*, young leaves fix a higher proportion of $^{14}CO_2$ directly through Rubisco, as compared to older leaves (Moore *et al.* 1986). Hence, establishment of Kranz anatomy limits young leaves of both monocots and dicots in conducting C_4 photosynthesis (Nelson & Dengler 1992). Whether such changes in C_4 enzyme activity occur during maturation of the dicot *C. gynandra* has not yet been tested.

In this work, two species of the genus *Cleome* were used, *Cleome gynandra* as a C_4 plant and *Cleome spinosa* as a C_3 plant (Marshall *et al.* 2007) to investigate the biochemistry of dicotyledonous NAD-ME type C_4 plants. The genus *Cleome* belongs to the Cleomaceae, which is phylogenetically close to the Brassicaceae model plant *Arabidopsis thaliana* (Hall *et al.* 2002; Brown *et al.* 2005; Inda *et al.* 2008). Our experiments addressed two major questions: is the C_4 pathway of the dicot *C. gynandra* modulated with age, and which enzymes play a role in NAD-ME type C_4 photosynthesis performed in the dicot *C. gynandra*?

MATERIAL AND METHODS

Plant growth

Cleome gynandra, *C. spinosa* and *Nicotiana benthamiana* plants were grown in soil in a glasshouse. Natural sunlight was supplemented by artificial 400 W spotlights mounted ~1.5 m above the ground level from 05:30 to 09:30 h to extend the daily light period to 16 h. Natural illumination was limited to 60,000 lux using shading. The temperature was kept constant at 24 °C through artificial ventilation. The humidity was measured between 75% and 90%.

Harvesting leaf material

Leaf material was harvested from mature green leaves, when not used for leaf age gradient experiments. For all experiments, leaves were snap-frozen in liquid nitrogen immediately after harvesting. For age gradient experiments, plants were grown for 4 weeks, at which point eight to ten true leaves were developed. Material from ten plants was pooled. The second youngest leaf was taken as the young leaf sample, the fifth youngest leaf was taken as the mid-aged leaf sample and the second oldest leaf was taken as the old leaf sample. Leaves were harvested after 6 h of illumination (at 11:30 h).

Enzyme assays

Leaves were cut at the base and immediately frozen in liquid nitrogen. Samples were ground to a fine powder under constant supply of liquid nitrogen, using a mortar and pestle. Leaves were harvested according to the described routine for the age gradient. Aliquots of 10 mg of leaf powder for each developmental stage were transferred into pre-chilled reaction tubes, and 1 ml of extraction buffer containing 25 mM Tris HCl (pH 7.5), 1 mM magnesium sulphate, 1 mM ethylenediaminetetraacetic acid, 5 mM DTT, 0.2 mM phenylmethylsulphonyl fluoride and 10% (v/v) glycerol were added to each sample and immediately vortexed to mix the ingredients before the extraction buffer froze. Tubes were thawed at room temperature for 10 min. After inverting the tubes four to six times, samples were centrifuged for 5 min at 4 °C and 16,000 *g*, and 500 µl of the supernatant were transferred into a fresh tube.

Enzyme activity measurements were conducted using coupled assays. PEPC activity was determined as described in Jiao & Chollet (1988). NAD-ME and NADP-ME were assayed as described in Hatch & Mau (1977). The protocol of Walker *et al.* (1995) was used to determine PEPCK activity in the reverse direction. The method of Hatch & Mau (1973) was performed to determine AspAT and AlaAT activity. MDH assays were performed according to Johnson & Hatch (1970). Leaf extracts of each developmental stage were desalted using size exclusion chromatography on NAP5 columns [Sephadex (GE Healthcare, Barrington, IL, USA) G-25 DNA Grade] prior to all enzyme assays.

Determination of chlorophyll and protein

Total chlorophyll was measured according to Porra (2002). Total protein determination was carried out using the BCA™ Protein Assay Kit (Thermo Scientific, Rockford, IL, USA). Protein concentration was determined in leaf extracts of each development stage according to manufacturer's instructions. Leaf extract was generated as described above.

Metabolite profiling

The extraction of leaves was performed according to Fiehn (2007) and GC/MS was performed according to Lee & Fiehn (2008). Analysis of metabolites was done *via* GC/MS-TOF [Autosampler: Gerstel, (Mülheim an der Ruhr, Germany), MPS2 XL; GC: Agilent Technologies (Santa Clara, CA, USA) 7890A; MS-TOF: Waters (Milford, MA, USA) GCT premier] and analysed using the MassLynx software package supplied with the instrument (Waters). Ribitol was added as an internal standard and the ratio of the area of each metabolite and the corresponding ribitol area was referred to as the ribitol response factor (RRF).

Native polyacrylamide gel electrophoresis

Aliquots of 50 mg of leaf powder were transferred into pre-chilled 1.5-ml reaction tubes, with 200 l l of extraction buffer (50 mm Tris HCl pH 6.8, 10% (v/v) glycerol, 5% 2-mercaptoethanol, 0.001% bromophenol blue) added to each sample. Samples were vortexed immediately to avoid freezing of the extraction buffer. Tubes were thawed for 10 min at room temperature and centrifuged at 16,000 g for 5 min at 4 °C. Supernatants were transferred into fresh tubes. Gels consisted of a stacking gel, which was ~1-cm wide, followed by a running gel, which was ~6-cm wide (Ferl et al. 1979). AspAT gels were stained with L-cysteine sulphinic acid as described in Stejskal (1994).

Transient expression of proteins in N. benthamiana

Cleome gynandra AspAT was cloned into pMDC83 via pDONR207. The fragment was amplified using primer A (GGGGACAAGTTTGTACAAAAAAGCAGGCTGAGAG ATGG CTATGGCTATGG) and primer B (GGGGACCACTTTGTAC AAGAAAGCTGG GTCAGACTTAGAGACCTCATGGATCGC). To detect the cellular localisation of proteins, vectors containing expression cassettes were transformed into N. benthamiana leaves via agrobacterial infection (Bendahmane et al. 1999). The co-infiltration with a mitochondrial reporter construct (ScCOX4: mCherry; Nelson et al. 2007) and the AspAT-containing construct was achieved by mixing liquid Agrobacterium cultures prior to infiltration (Walter et al. 2004). Localisation was observed in epidermal peels using epifluorescence microscopy with appropriate filters for mCherry and for GFP.

RESULTS

Activity of C_4 enzymes

To assess possible age-related differences in the mode of C_4, the activity of enzymes involved in C_4 photosynthetic pathways was determined in extracts of Cleome leaves of different age. The three decarboxylating enzymes NAD-ME, NADP-ME and PEPCK were tested, as well as PEPC, AspAT, alanine aminotransferase (AlaAT) and both NAD-dependent and NADP-dependent malate dehydrogenase (NAD-MDH and NADP-MDH).

In mid-aged mature leaves, the activity of PEPC was 1.8 U g^{-1} FW in the C_4 species and 0.02 U g^{-1} FW in the C_3 species. NAD-ME showed the highest activity among the decarboxylating enzymes (0.59 U g^{-1} FW) and exceeded the activity in C. spinosa (0.04 U g^{-1} FW) by an order of magnitude (Fig. 1A). Also, AspAT (1.51 U g^{-1} FW) and AlaAT (2.89 U g^{-1} FW) in mid-aged C. gynandra leaves were significantly ($P < 0.05$) higher than the corresponding activities in C. spinosa leaves (0.53 and 0.86 U g^{-1} FW, respectively). There was no significant difference between C. gynandra and C. spinosa in the activity of NAD-MDH, which is also part of the C_4 shuttling pathway in NAD-ME-dependent C_4 plants. Mid-aged leaves of C. gynandra showed substantial activity of PEPCK, reaching two-thirds of the NAD-ME activity in the same sample (Fig. 1). NADP-ME showed low activity in mid-aged leaves of C. gynandra (0.03 U g^{-1} FW) when compared to the other decarboxylating enzymes. In mid-aged leaves of C. spinosa, NADP-ME activity was significantly higher (0.26 U g^{-1} FW).

It was then tested whether differences in enzyme activity exist in different aged leaves. In old leaves of C. gynandra, the NAD-ME activity (0.28 U g^{-1} FW) decreased significantly ($P < 0.05$) by 50% when compared to the mid-aged leaf sample (Fig. 1; Table 1). Also the AlaAT activity in old C. gynandra leaves decreased significantly (Table 1). In contrast, the PEPCK activity increased by 30% in old C. gynandra leaf samples and the AspAT activity increased significantly ($P < 0.05$) by 60% (Table 1). PEPC and both NAD- and NADP-dependent MDHs showed only minor changes in activity in old as compared to mid-aged leaf samples. NADP-ME activity increased by 40% (Table 1), but NADP-ME was still the least active decarboxylating enzyme in C. gynandra (Fig. 1). Old leaves of C. spinosa showed a general age-related reduction in enzyme activities (Fig. 1). NADP-ME showed lower activity compared to mid-aged leaves. Also, both NAD- and NADP-dependent MDH showed significantly decreased activity (Fig. 1).

In young C. gynandra leaves, PEPC showed 50% less activity when compared to mid-aged leaf samples. PEPC activity

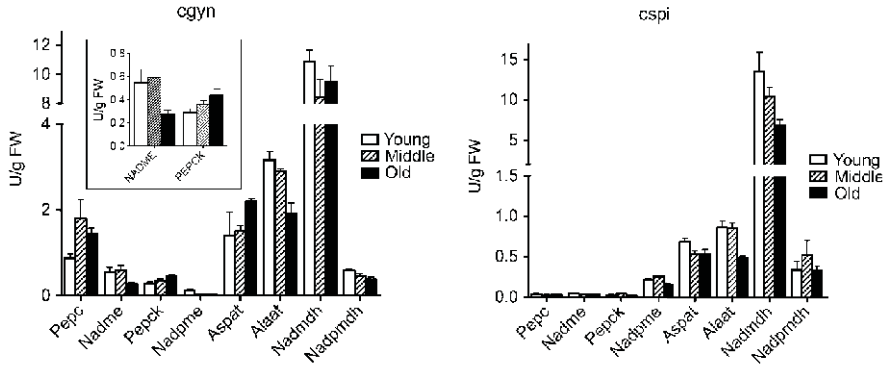


Fig. 1. Enzyme activity in young, mid-aged and old leaves. Means and standard errors of n = 5 are plotted. cgyn C. gynandra; cspi C. spinosa; pepc phosphoenolpyruvate carboxylase; nadme, NAD-dependent malic enzyme; pepck PEP carboxykinase; nadpme, NADP-dependent malic enzyme; aspat, aspartate aminotransferase; alaat, alanine aminotransferase; nadmdh, NAD-dependent malate dehydrogenase; nadpmdh, NADP-dependent malate dehydrogenase.

Table 1. Age-related fold-change of enzyme activity in C. gynandra leaves.

fold change	young/middle	middle/old	young/old
PEP carboxylase	2.0	0.9	1.8*
NAD malic enzyme	1.3	0.5*	0.6
PEP carboxykinase	1.3	1.3	1.7*
NADP malic enzyme	0.4	1.4	0.7
aspartate aminotransferase	1.6	1.6**	2.2
alanine aminotransferase	0.9	0.7**	0.6**
NAD malate dehydrogenase	0.8	1.1	0.9
NADP malate dehydrogenase	0.8	0.9	0.6*

Asterisks mark significant fold changes. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

in young leaves was significantly lower than in old leaves. Also, the AspAT activity in young C. gynandra leaves was 40% lower as compared to mid-aged leaves, albeit this difference was not statistically significant. The activities of other enzymes showed only minor changes. In young C. spinosa leaves, the only significantly different activity compared to mid-aged leaves was observed for AspAT, which showed increased activity in young leaves.

Although NAD-ME activity was reduced significantly from mid-aged to old leaves, the sum of activities of all decarboxylating enzymes (NAD-ME, NADP-ME and PEPCK) did not change significantly between leaves of different ages.

Total chlorophyll and protein content

Ratios of total protein per total chlorophyll and total protein per fresh weight in C. gynandra and C. spinosa were compared. C. spinosa contained significantly more protein than C. gynandra, both on the basis of chlorophyll and fresh weight in mid-aged leaves (Fig. 2). No significant difference was detected between the C_3 and the C_4 plant in young and old leaves.

Metabolite analysis

Metabolites were analysed to obtain a more detailed view of the differences between C_3 and C_4 metabolism. In total, 49 distinct metabolites were analysed, of which 31 could be detected and quantified reliably.

In young leaf samples, nine of 31 metabolites were more abundant in C. gynandra; this number increased to 24 in old leaves. The main metabolites of the C_4 shuttling pathway, e.g., alanine, aspartate and malate, were significantly more abundant in mid-aged and old leaves of C. gynandra, as compared to C. spinosa (Fig. 3B,C). However, no significant difference was detected for aspartate and malate in young leaves (Fig. 3A). Also, asparagine, which can be synthesised from aspartate by aspartate-ammonia ligase, fumarate, which can be synthesised from malate by fumarate hydratase, and glutamate, which is a central metabolite of nitrogen metabolism, are significantly more abundant in mid-aged and old C. gynandra leaves compared to levels in C. spinosa. However, there was no significant difference in these metabolites in young leaves (Fig. 3). The contents of the sugars glucose, fructose, mannose, xylose and sucrose did not differ; only maltose showed a significantly higher abundance in C. spinosa in all leaf ages (Fig. 3). Xylose abundance appeared to be greater in young C. spinosa samples. Sucrose abundance was almost identical when comparing C. gynandra and C. spinosa leaves of the same age (Fig. 3). Among the 11 amino acids that could be reliably detected and quantified, aspartate, alanine and asparagine were more abundant in C_4 samples, whilst others were detected in higher concentrations in the C_3 plant, e.g., methionine and leucine.

Aspartate aminotransferase in C. gynandra

After analysing age-related differences in enzyme activity and steady state metabolite pools, AspAT function was analysed in detail in mature mid-aged leaves. Transcriptome analysis detected five AspAT isoforms in leaf transcriptomes of both C. gynandra and C. spinosa (Bräutigam & Weber 2011a,b). Native protein gels followed by activity staining were performed to determine the numbers of isozymes and their individual activities in these two species. For both species, five distinct bands stained for AspAT activity (Fig. 4). In C. gynandra, the five bands were spread wider as compared to in C. spinosa. The C. gynandra band migrating furthest into the gel showed the highest activity of all bands, as evidenced by the intensity of staining. The four other bands of C. gynandra and the five bands of C. spinosa showed comparable activity.

Subcellular localisation of the highly abundant AspAT isoform of C. gynandra

Combining quantitative (Bräutigam et al. 2011a) and qualitative (Bräutigam et al. 2011b) next generation sequencing information, as proposed in Bräutigam & Gowik (2010), identified the sequence of the upregulated CgAspAT. A CgAspAT-GFP fusion construct based on the full-length contig of

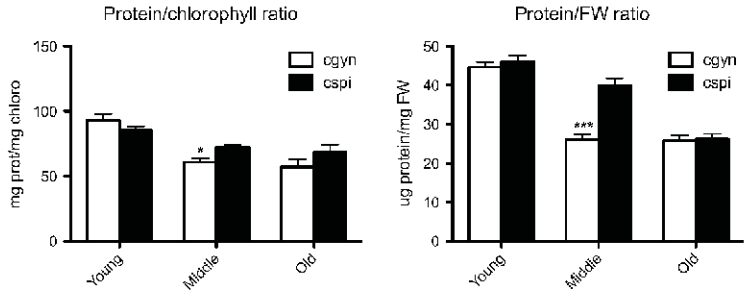


Fig. 2. Total chlorophyll and protein in a C_3 and a C_4 plant. Significant changes are marked with asterisks, * $P < 0.05$; *** $P < 0.001$.

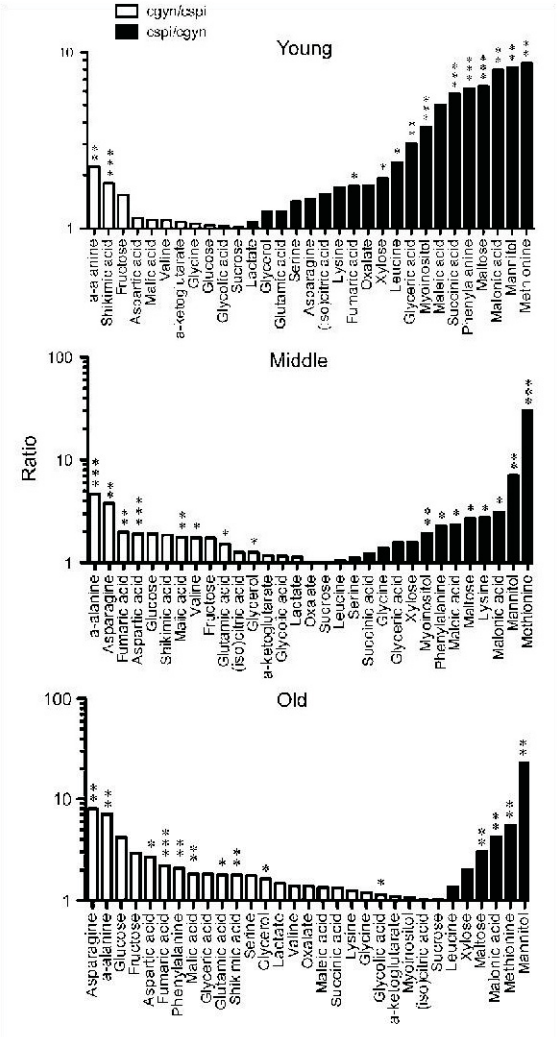


Fig. 3. Metabolite analysis in three leaf age stages. The ratios between C_3 and C_4 leaves are plotted. Significant changes were calculated with Student's t-tests (n = 5) and are marked with asterisks, *P < 0.05; **P < 0.01; ***P < 0.001.

this upregulated AspAT was created to investigate the subcellular localisation of the transcriptionally abundant isozyme of AspAT (Bräutigam et al. 2011a) in N. benthamiana, which maps to the mitochondrial AspAT1 of A. thaliana. Mitochondria were tagged with a fluorescent marker as a positive control (Nelson et al. 2007). The CgAspAT::GFP fusion protein was detected as small dots, both in photosynthetic cells as well as epidermal cells of N. benthamiana leaves. Co-localisation was determined in epidermal cells to minimise interference from chlorophyll fluorescence. The organelles labelled with GFP were detected on the periphery of the cell (Fig. 5B). Structures in the same position were labelled with the mitochondrial marker (Fig. 5A) and the overlay revealed perfect congruence of the fluorescence signals (Fig. 5D). Chloroplasts

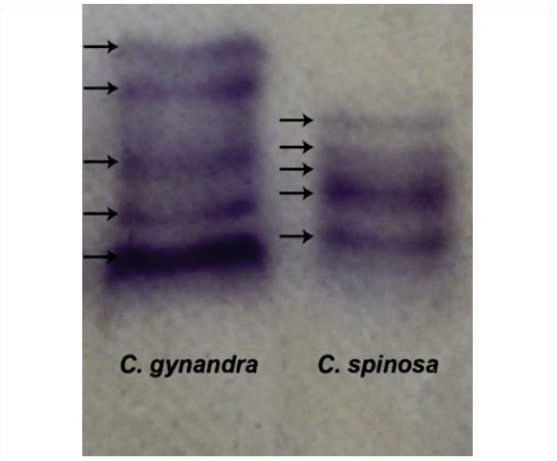


Fig. 4. Isozyme analysis of AspATs in C. gynandra and C. spinosa. Arrowheads point to isozymes.

in the sample (Fig. 5C) caused minor signals in the mCherry channel for the mitochondrial marker (Fig. 5A, C).

DISCUSSION

Leaf age-related changes in C_4 photosynthesis in C. gynandra

Photosynthetic C_3 and C_4 plant tissues have distinct biochemical properties that can be assessed using enzyme activity assays. In Cleomaceae, C. gynandra performs C_4 photosynthesis of the NAD-ME type, whereas C. spinosa performs C_3 photosynthesis (Marshall et al. 2007). Enzymatic tests were performed to qualify and quantify the activity of C_4 marker enzymes, including PEPC and the main decarboxylating enzyme NAD-ME, and to provide information about establishment of the C_4 syndrome during leaf development. In addition, metabolite analysis was performed via GC/MS-TOF to assess the steady state metabolite pools associated with the photosynthetic subtypes in C. gynandra and C. spinosa.

The soluble sugar contents in the C_4 and the C_3 species were comparable. Sucrose was found at nearly the same concentration in both species (C. gynandra/C. spinosa ratio, young: 1.0, mid: 1.0, old: 1.0). Only maltose, which is exported from the chloroplast following metabolic conversion of transitory starch at night (Weise et al. 2004), showed consistently higher abundance in C. spinosa. Both C_4-related metabolite pools as well as C_4-related enzyme activities, however, clearly differed between the two species. NAD-ME had highest activity among the decarboxylating enzymes in mid-aged leaves of C. gynandra, which confirms C. gynandra as an NAD-ME type plant (Marshall et al. 2007). The ratios between NAD-ME and NADP-ME (18:1) and between NAD-ME and PEPCK (2:1) were also comparable to earlier results (Marshall et al. 2007; Bräutigam et al. 2011a). Minor activities of secondary decarboxylating enzymes in C_4 plants or general decarboxylating enzymes in C_3 plants are probably related to their function in housekeeping processes, such as amino acid biosyntheses (Brown et al. 2010), or in other

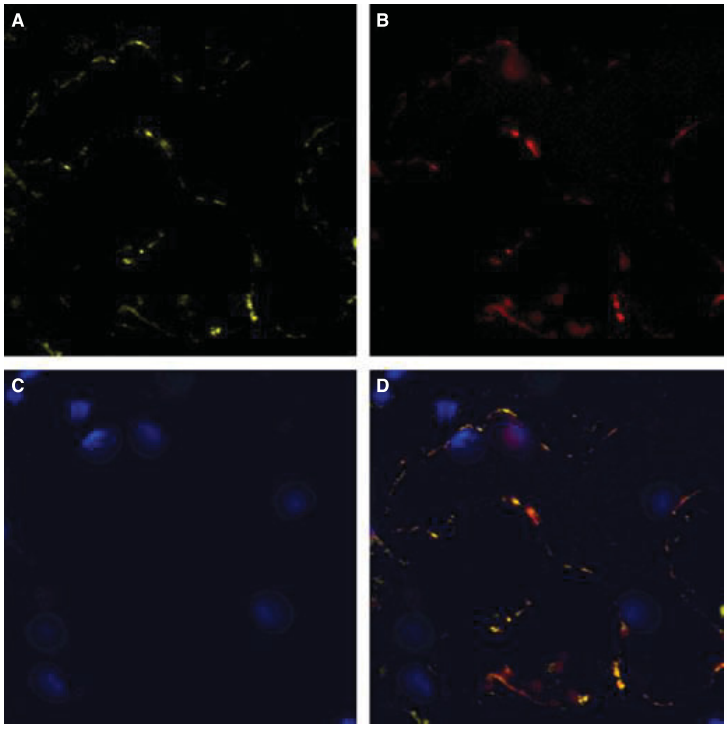


Fig. 5. AspAT::GFP co-localised with the mitochondrial marker. A mCherry fluorescence of the mitochondrial reporter construct; B GFP fluorescence of the AspAT::GFP fusion construct; C chloroplast autofluorescence; D merged image of A–C.

processes, such as defence (Maurino *et al.* 2001). Intermediates of the NAD-ME type C_4 pathway, namely aspartate and malate, showed higher ribitol response factors (RRFs; see Material and Methods) in mid-aged *C. gynandra* leaves (aspartate: 1.9-fold, malate: 1.8-fold, alanine: 4.7-fold). The high abundances of intermediates of NAD-ME type C_4 photosynthesis is in agreement with predominant activity of enzymes of the NAD-ME cycle. Both MDH activities did not significantly increase in *C. gynandra* (Fig. 1; Table 1) when comparing mid-aged leaf samples, which is surprising, given that the expected increase in MDH activity is necessary for proper C_4 carbon shuttling in *C. gynandra*. However, the activity of NAD-MDH was already very high in the C_3 species (Fig. 1). This may indicate that the MDH activity is high enough to support C_4 photosynthesis in addition to its housekeeping function. Alternatively, since total activity was measured, mitochondrial activity may have been higher, which may have been offset by lower peroxisomal, cytosolic and/or plastidic activities. Transcriptome analysis of *C. gynandra* and *C. spinosa* did not reveal an increase in mitochondrial NAD-MDH (Bräutigam *et al.* 2011a), which favours the first hypothesis. Although the activity of NAD-MDHs was steady, the pool size of malate was larger in the C_4 species.

The pattern of C_4-related enzyme activities changes with leaf age. Analysis of the decarboxylating enzymes revealed that *C. gynandra* NAD-ME activity decreased in old leaves by 50%, while PEPCK activity was increased significantly compared to that of young leaves and exceeded the NAD-ME activity by 50%. These results are unexpected for an NAD-ME type plant, since previous data indicated only minor PEPCK activity in NAD-ME plants (Hatch 1987). Significant PEPCK activity in leaves is in contrast to a recent report, which includes immunostaining of *C. gynandra* crude leaf extract (Voznesenskaya *et al.* 2007); no binding of PEPCK-specific antibody to the extract was detected in this experiment. In this previous study, the developmental stage of harvested leaves was referred to as mature. Possibly, the leaf age was younger compared to what we call mid-age. Only PEPCK type C_4 plants are known to show major secondary decarboxylating enzyme activity for NAD-ME (Hatch *et al.* 1988). The observed gradient in decarboxylating enzyme activity can be explained by the hypothesis that *C. gynandra* gradually shifts its C_4 pathway during leaf development from NAD-ME type to a mixed NAD-ME and PEPCK type. If this hypothesis holds, the activity of other enzymes involved in the C_4 shuttling pathway ought to change as well, namely less activity of NAD-ME type specific enzymes and constant activity of enzymes that are required for both types of C_4 photosynthesis. Indeed, AlaAT, which is required for NAD-ME type C_4 photosynthesis but has no use in the PEPCK pathway, showed significantly lower activity in old *C. gynandra* leaves when compared to mid-aged leaves. The shift from NAD-ME type to PEPCK type may decrease the demand for this enzyme. At the same time, PEPC, which is required for all types of C_4 photosynthesis, did not show decreased activity when comparing old and mid-aged leaves of *C. gynandra*. However, the pool size of intermediates in the NAD-ME pathway further increases in old C_4 leaves compared to those in the C_3 plant. The C_4/C_3 ratio of alanine increased from 4.7-fold in mid-aged leaves to 7.0 in old leaves. This indicates that although the decarboxylating activity is at least in part taken over by PEP-CK, the C_3 amino acid alanine remains

important. If aspartate continues to be shuttled for a PEP-CK type of cycle and PEP is regenerated, the amino group ratio between mesophyll and bundle sheath becomes unbalanced. Possibly, glutamate (increased significantly in the C_4 species in mid-aged and old leaves) and 2-oxoglutarate (not significantly increased) play a role in the C_4 cycle (Fig. 3). A shift from NADP-ME type to a mixed NADP-ME and PEPCK type has been observed in old leaves for the monocot plant *Zea mays* (Wingler *et al.* 1999). Our analysis of protein to chlorophyll and protein to fresh weight ratios supports the notion that major changes occur from young to mid-aged old leaves, which altered the protein/chlorophyll and protein/fresh weight ratios observed in mature mid-aged leaves (Fig. 2). The significantly lower protein content in mid-aged C_4 leaves is most likely due to the efficiency of the C_4 pathway. It has been shown that C_4 plants contain less Rubisco protein compared to C_3 plants (Ku *et al.* 1979). Transcriptome analysis of *Cleome* species also revealed that functional clusters of proteins, which are directly or indirectly involved in photosynthesis, are differentially expressed in C_4 photosynthesis (Bräutigam *et al.* 2011a). That is, transcripts encoding Rubisco subunits and enzymes of photorespiration are expressed at lower levels in the C_4 plant. In addition, it was found that genes encoding components of cytosolic and plastidial ribosomes were also expressed at lower levels, which likely also contributes to a lower protein content in these leaves. Further, it was found in a recent study using C_3, C_3-C_4 intermediate and C_4 species of *Flaveria* that the C_4 species always displayed a higher carbon-to-nitrogen ratio than the C_3 species, which is indicative of a lower protein content in the C_4 species (Gowik *et al.* 2011). Thus, since changes in total leaf protein concentration are apparently a consequence of functional adaptations in C_4 photosynthesis rather than natural variation, total protein content was not used in this study to normalise enzyme activity data.

The PEPC activity in young leaves was low compared to that in other leaf ages (Fig. 1). This indicates that C_4 carbon fixation is not yet fully expressed at this developmental stage. Also, the C_4/C_3 ratio of metabolites involved in the NAD-ME pathway, *i.e.*, aspartate (young: 1.1-fold, mid: 1.9-fold), malate (young: 1.1-fold, mid: 1.8-fold) and alanine (young: 2.2-fold, mid: 4.7-fold), is lower in young compared to mid-aged leaves. This finding is consistent with the hypothesis that C_4 photosynthesis is not fully developed in young *C. gynandra* leaves compared to mid-aged and old leaves. Instead, young leaves may use C_3 type photosynthesis or a mix of C_4 and C_3 photosynthesis. It is also possible that young leaves perform less total photosynthesis. Based on the protein ratios, major changes occur when leaves mature and fully implement C_4 photosynthesis. However, the current results do not allow the rejection of either hypothesis.

AspAT isozymes

The aspartate aminotransferases (AspATs) were investigated in detail in mid-aged *Cleome* leaves. They convert OAA to aspartate and *vice versa*. The canonical model for NAD-ME type C_4 plants requires two isoforms of this enzyme, a cytosolic isoform in mesophyll cells and a mitochondrial isoform in bundle sheath cells (Taniguchi & Sugiyama 1990; Taniguchi *et al.* 1995). This model was based on activity gels of *P. miliaceum* (Hatch & Mau 1973). mRNA-Seq data (Bräutigam *et al.* 2011a), however, indicate that in *C. gynandra* only one isoform is upregulated at the transcriptional level, which contradicts the canonical model.

In both *C. gynandra* and *C. spinosa*, a total of five AspAT isozymes were detected in native enzyme activity gels (Fig. 4), which is in accordance with the number of different cDNAs detected at the transcriptome level (Bräutigam *et al.* 2011a). This number is also identical to the number of AspAT isozymes encoded by the genome of *A. thaliana* (Schultz & Coruzzi 1995; Wilkie *et al.* 1995). Only one of these five isozymes showed very high activity in *C. gynandra* compared to any of the other isozymes in both species (Fig. 4). This result is in agreement with findings of mRNA-Seq, where only a single upregulated AspAT transcript was detected (Bräutigam *et al.* 2011a). In models of NAD-ME type C_4 photosynthesis, two reactions catalysed by AspAT take place, one of them in the cytosol and one in mitochondria. Usually proteins are targeted to only one location in the cell; however, some proteins are present in multiple locations due to dual targeting (Small *et al.* 1998). Subcellular localisation studies of CgAspAT1 were performed to assess the localisation of the highly abundant AspAT isoform in *C. gynandra*. The overlay of the signals of GFP-tagged CgAspAT with a mitochondrial-localised control fused to the mCherry protein confirmed co-localisation of both gene products. Hence, the highly abundant AspAT isozyme is targeted to mitochondria. Multiple targeting was excluded since no GFP signal was detected in other compartments of the cell or in the cytosol. In *P. miliaceum* NAD-ME type C_4 photosynthesis is concomitant with increased activity of two different AspAT isoforms in bundle sheath and mesophyll cells (Hatch & Mau 1973). Since mesophyll and bundle sheath cells cannot be easily separated in dicots, in contrast to monocots, only whole leaf extracts were examined (Fig. 4). However, in *P. miliaceum*, even in whole leaf extracts the change in activity of two isoforms, not just one, was apparent on activity gels of whole leaf extracts. It is therefore concluded that CgAspAT1 performs both AspAT-catalysed reactions of the C_4 pathway in mitochondria, which indicates that the NAD-ME C_4 pathway in *C. gynandra* differs from that of *P. miliaceum*.

The mitochondrial localisation of AspAT requires additional transport processes across the mitochondrial membrane. The substrates of the AspAT reaction, namely OAA and glutamate, have to be imported, whereas the products, aspartate and 2-oxoglutarate, have to be exported. Despite detailed knowledge on soluble proteins in C_4 photosynthesis, the adaption of membrane proteins remains largely unknown (Bräutigam *et al.* 2008; Weber & von Caemmerer 2010; Bräutigam & Weber 2011a; b; Furumoto *et al.* 2011). Since it is known that proteins involved in C_4 photosynthesis are modified in localisation and abundance rather than newly invented (Sage 2004), it is assumed that transporters that already exist in C_3 plants are re-dedicated to C_4-related transport processes in *C. gynandra* (Furumoto *et al.* 2011). Two transporters were characterised biochemically in mitochondria of *Vigna sinensis* and *Pisum sativum*; one catalyses the counter-exchange of 2-oxoglutarate and OAA (Desantis *et al.* 1976) and the other catalyses the counter-exchange of aspartate and glutamate (Vivekananda & Oliver 1989). Taken together, both carriers are able to shuttle all required sub-

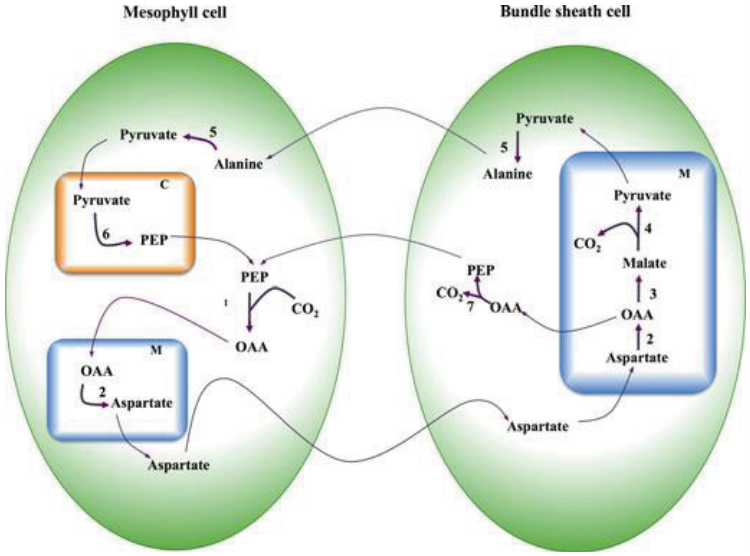


Fig. 6. New model of the NAD-ME pathway. Thick arrows indicate reaction, thin arrows indicate transport process. M=mitochondrium. C=chloroplast. 1 = PEPC, 2 = AspAT, 3 = MDH, 4 = NAD-ME, 5 = AlaAT, 6 = PPDK, 7 = PEPCK.

strates and products for AspAT-catalysed transamination, and this balances the amino groups between the cytosol and the mitochondria. It is currently not known whether and how the transport capacities of mitochondria are adjusted in NAD-ME plants. In evolutionary terms, establishing the *P. miliaceum* model requires the upregulation of two different genes and the restriction of one gene to the mesophyll and of a second gene to bundle sheath cells. Establishing the *C. gynandra* model requires the general increase in expression of one gene and the up-regulation of two transport proteins, if transport capacity is indeed limiting. mRNA-Seq data showed increased expression levels of only three mitochondrial transport proteins in *C. gynandra*, namely the dicarboxylate carriers DIC1 and DIC2, and a phosphate transporter (Bräutigam *et al.* 2011a). Neither of the dicarboxylate carriers shows transport activity for aspartate or any other amino acid in the C_3 plant or the close relative *A. thaliana* (Palmieri *et al.* 2008). The phosphate transporter is a *bona fide* phosphate transporter in grasses (Takabatake *et al.* 1999). It remains unclear whether one of the DIC transporters can transport aspartate in C_4 plants. In summary, it is proposed that *C. gynandra* performs a modified C_4 cycle in which not only the AspAT-catalysed reaction occurs in bundle sheath cells, but also in the mitochondria of mesophyll cells. Required transporters were not found in this work, but the required transport processes can be performed by equivalents of at least partially described transporters from C_3 plant mitochondria. The evolutionary changes required are not exceptionally larger than those required for the *P. miliaceum* model.

A model of NAD-ME in the dicot *C. gynandra*

This work shows that *C. gynandra* does not perform NAD-ME type C_4 photosynthesis exclusively, but also uses PEPCK as a decarboxylation enzyme. PEPCK use is more pronounced in older leaves, while very young leaves show only limited PEPC activity. An increase in PEPC in mid-aged leaves suggests that C_4 photosynthesis may be used predominantly in more mature leaves, while young leaves use either C_3, a mix of C_3 and C_4 photosynthesis or perform less total photosynthesis. Based on the results of age-dependent analyses, it is proposed that *C. gynandra* passes through key changes in the carbon assimilation mechanism during leaf development, namely showing a shift from C_3 photosynthesis or a mix of C_3 and C_4 photosynthesis in young leaves to exclusively, or almost exclusively, C_4 photosynthesis in old leaves, and at the same time, a shift from NAD-ME C_4 type in young leaves to a combination of NAD-ME and PEPCK C_4 type in old leaves. The decarboxylation enzymes are indeed flexible, although plasticity based on developmental changes rather than environmental changes as proposed in (Furbank 2011) was demonstrated.

In addition, the presence of only a single up-regulated isoform of AspAT, which is localised to mitochondria, leads to a new model for C_4 photosynthesis in mesophyll cells. Instead of a cytosolic conversion of OAA to aspartate, the reaction takes place in mitochondria. Consequently, a new model for NAD-ME type C_4 photosynthesis is proposed for *C. gynandra* (Fig. 6).

REFERENCES

Bassi R., Passera C. (1982) Effect of growth conditions on carboxylating enzymes of *Zea mays* plants. *Photosynthesis Research*, **3**, 53–58.

Bendahmane A., Kanyuka K., Baulcombe D.C. (1999) The Rx gene from potato controls separate virus resistance and cell death responses. *Plant Cell*, **11**, 781–792.

Black C.C. (1973) Photosynthetic carbon fixation in relation to net CO_2 uptake. *Annual Review of Plant Physiology and Plant Molecular Biology*, **24**, 253–286.

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

Bräutigam A., Weber A.P.M. (2011a) Transport processes – connecting the reactions of C_4 photosynthesis. In: Agepati S., Sage Rowan F. (Eds), *Advances in Photosynthesis and Respiration*. Vol. 32, Raghavendra. Springer, Berlin, pp 199–219.

Bräutigam A., Weber A.P.M. (2011b) Do transport process limit photosynthesis? *Plant Physiology*, **155**, 43–48.

Bräutigam A., Hofmann-Benning S., Weber A.P.M. (2008) Comparative proteomics of chloroplast envelopes from C_3 and C_4 plants reveals specific adaptations of

the plastid envelope to C_4 photosynthesis and candidate proteins required for maintaining C_4 metabolite fluxes. *Plant Physiology*, **148**, 568–579.

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., Weber A.P.M. (2011a) An mRNA blueprint for C_4 photosynthesis derived from comparative transcriptomics of closely related C_3 and C_4 species. *Plant Physiology*, **155**, 142–156.

Bräutigam A., Mullick T., Schliesky S., Weber A.P.M. (2011b) Critical assessment of assembly strategies for non-model species mRNA-Seq data and application of next-generation sequencing to the comparison of C_3 and C_4 species. *Journal of Experimental Botany*, **62**, 3093–3102.

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

Brown N.J., Palmer B.G., Stanley S., Hajaji H., Janacek S.H., Astley H.M., Parsley K., Kajala K., Quick W.P., Trenkamp S., Fernie A.R., Maurino V.G., Hibberd J.M. (2010) C_4 acid decarboxylases required for C_4 photosynthesis are active in the mid-vein of the C_3 species *Arabidopsis thaliana*, and are important in sugar and amino acid metabolism. *Plant Journal*, **61**, 122–133.

Desantis A., Arrigoni O., Palmieri F. (1976) Carrier-mediated transport of metabolites in purified bean mitochondria. *Plant and Cell Physiology*, **17**, 1221–1233.

Ferl R.J., Dlouhy S.R., Schwartz D. (1979) Analysis of maize alcohol-dehydrogenase by native-SDS 2-dimensional electrophoresis and auto-radiography. *Molecular and General Genetics*, **169**, 7–12.

Fiehn O. (2007) Validated high quality automated metabolome analysis of *Arabidopsis thaliana* leaf disks. In: Nikolau B.J., Wurtele E.S. (Eds), *Concepts in Plant Metabolomics*. Springer, Berlin, pp 1–18.

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

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.M., Izui K. (2011) A plastidial sodium-dependent pyruvate transporter. *Nature*, **476**, 472–475.

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

Hall J.C., Sytsma K.J., Iltis H.H. (2002) Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data. *American Journal of Botany*, **89**, 1826–1842.

Hatch M.D. (1987) C_4 photosynthesis – a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochimica et Biophysica Acta*, **895**, 81–106.

Hatch M.D., Mau S.L. (1973) Activity, location and role of aspartate aminotransferase and alanine aminotransferase isoenzymes in leaves with C_4 pathway photosynthesis. *Archives of Biochemistry and Biophysics*, **156**, 195–206.

Hatch M.D., Mau S.L. (1977) Association of NADP-linked and NAD-linked malic enzyme activities in *Zea Mays* – relation to C_4 pathway photosynthesis. *Archives of Biochemistry and Biophysics*, **179**, 361–369.

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

Hatch M.D., Agostino A., Burnell J.N. (1988) Photosynthesis in phosphoenolpyruvate carboxykinase-type C_4 plants – activity and role of mitochondria in bundle sheath-cells. *Archives of Biochemistry and Biophysics*, **261**, 357–367.

Inda L.A., Torrecilla P., Catalán P., Ruiz-Zapata T. (2008) Phylogeny of *Cleome L.* and its close relatives *Podandrogyne ducke* and *Polanisia raf.* (Cleomoideae, Cleomaceae) based on analysis of nuclear ITS sequences and morphology. *Plant Systematics and Evolution*, **274**, 111–126.

Jiao J.A., Chollet R. (1988) Light dark regulation of maize leaf phosphoenolpyruvate carboxylase by in vivo phosphorylation. *Archives of Biochemistry and Biophysics*, **261**, 409–417.

Johnson H.S., Hatch M.D. (1970) Properties and regulation of leaf nicotinamide-adenine dinucleotide phosphate-malate dehydrogenase and malic enzyme in plants with C_4 dicarboxylic acid pathway of photosynthesis. *Biochemical Journal*, **119**, 273–280.

Ku M.S.B., Schmitt M.R., Edwards G.E. (1979) Quantitative determination of RuBP carboxylase-oxygenase protein in leaves of several C_3 and C_4 plants. *Journal of Experimental Botany*, **30**, 89–98.

Lee D.Y., Fiehn O. (2008) High quality metabolomic data for *Chlamydomonas reinhardtii*. *Plant Methods*, 4, 7.

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

Maurino V.G., Saigo M., Andreo C.S., Drincovich M.F. (2001) Non-photosynthetic 'malic enzyme' from maize: a constitutively expressed enzyme that responds to plant defence inducers. *Plant Molecular Biology*, **45**, 409–420.

Miranda V., Baker N.R., Long S.P. (1981) Anatomical variation along the length of the *Zea mays* Leaf in relation to photosynthesis. *New Phytologist*, **88**, 595–605.

Moore B.D., Cheng S.H., Edwards G.E. (1986) The influence of leaf development on the expression of C-4 metabolism *in Flaveria trinervia*, a C-4 dicot. *Plant and Cell Physiology*, **27**, 1159–1167.

Nelson T., Dengler N.G. (1992) Photosynthetic tissue differentiation in C_4 plants. *International Journal of Plant Sciences*, **153**, S93–S105.

Nelson B.K., Cai X., Nebenfuhr A. (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. *Plant Journal*, **51**, 1126–1136.

Palmieri L., Picault N., Arrigoni R., Besin E., Palmieri F., Hodges M. (2008) Molecular identification of three *Arabidopsis thaliana* mitochondrial dicarboxylate carrier isoforms: organ distribution, bacterial expression, reconstitution into liposomes and functional characterization. *Biochemical Journal*, **410**, 621–629.

Porra R.J. (2002) The chequered history of the development and use of simultaneous equations for the accurate determination of chlorophylls a and b. *Photosynthesis Research*, **73**, 149–156.

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

Sage R.F., Christin P.A., Edwards E.J. (2011) The C_4 plant lineages of planet earth. *Journal of Experimental Botany*, **62**, 3155–3169.

Schultz C.J., Coruzzi G.M. (1995) The aspartate-aminotransferase gene family of Arabidopsis encodes isoenzymes localized to three distinct subcellular compartments. *Plant Journal*, **7**, 61–75.

Small I., Wintz H., Akashi K., Mireau H. (1998) Two birds with one stone: genes that encode products targeted to two or more compartments. *Plant Molecular Biology*, **38**, 265–277.

Stejskal J. (1994) Aspartate-aminotransferase isoenzymes in plants – comparison of two staining methods in polyacrylamide gels. *Biologia Plantarum*, **36**, 359–364.

Takabatake R., Hata S., Taniguchi M., Kouchi H., Sugiyama T., Izui K. (1999) Isolation and characterization of cDNAs encoding mitochondrial phosphate transporters in soybean, maize, rice, and *Arabidopsis*. *Plant Molecular Biology*, **40**, 479–486.

Taniguchi M., Sugiyama T. (1990) Aspartate-aminotransferase from *Eleusine coracana*, a C_4 plant – purification, characterization, and preparation of antibody. *Archives of Biochemistry and Biophysics*, **282**, 427–432.

Taniguchi M., Kobe A., Kato M., Sugiyama T. (1995) Aspartate-aminotransferase isozymes in *Panicum miliaceum* L, an NAD-malic enzyme-Type C_4 plant – comparison of enzymatic properties, primary structures, and expression patterns. *Archives of Biochemistry and Biophysics*, **318**, 295–306.

Vivekananda J., Oliver D.J. (1989) Isolation and partial characterization of the glutamate aspartate transporter from pea leaf mitochondria using a specific monoclonal antibody. *Plant Physiology*, **91**, 272–277.

Voznesenskaya E.V., Koteyeva N.K., Chuong S.D.X., Ivanova A.N., Barroca J., Craven L.A., Edwards G.E. (2007) Physiological, anatomical and biochemical characterisation of photosynthetic types in genus *Cleome* (Cleomaceae). *Functional Plant Biology*, **34**, 247–267.

Walker R.P., Trevanion S.J., Leegood R.C. (1995) Phosphoenolpyruvate carboxykinase from higher-plants – purification from cucumber and evidence of rapid proteolytic cleavage in extracts from a range of plant tissues. *Planta*, **196**, 58–63.

Walter M., Chaban C., Schutze K., Batistic O., Weckermann K., Nake C., Blazevic D., Grefen C., Schumacher K., Oecking C., Harter K., Kudla J. (2004) Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant Journal*, **40**, 428–438.

Weber A.P.M., von Caemmerer S. (2010) Plastid transport and metabolism of C_3 and C_4 plants – comparative analysis and possible biotechnological exploitation. *Current Opinion in Plant Biology*, **13**, 256–264.

Weise S.E., Weber A.P.M., Sharkey T.D. (2004) Maltose is the major form of carbon exported from the chloroplast at night. *Planta*, **218**, 474–482.

Wilkie S.E., Roper J.M., Smith A.G., Warren M.J. (1995) Isolation, characterization and expression of a cDNA clone encoding plastid aspartate-aminotransferase from *Arabidopsis thaliana*. *Plant Molecular Biology*, **27**, 1227–1233.

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

Manuscript 2

Comparative Transcriptome Atlases Reveal Altered Gene Expression Modules between Two Cleomaceae C_3 and C_4 Plant Species

This article is a *Plant Cell* Advance Online Publication. The date of its first appearance online is the official date of publication. The article has been edited and the authors have corrected proofs, but minor changes could be made before the final version is published. Posting this version online reduces the time to publication by several weeks.

Comparative Transcriptome Atlases Reveal Altered Gene Expression Modules between Two Cleomaceae C_3 and C_4 Plant Species[C][W][OPEN]

Canan Külahoglu,[a,1] Alisandra K. Denton,[a,1] Manuel Sommer,[a] Janina Maß,[b] Simon Schliesky,[a] Thomas J. Wrobel,[a] Barbara Berckmans,[c] Elsa Gongora-Castillo,[d] C. Robin Buell,[d] Rüdiger Simon,[c] Lieven De Veylder,[e,f] Andrea Bräutigam,[a,1] and Andreas P.M. Weber[a,2]

[a] Institute of Plant Biochemistry, Cluster of Excellence on Plant Sciences, Heinrich-Heine-University, 40225 Düsseldorf, Germany
[b] Institute of Informatics, Cluster of Excellence on Plant Sciences, Heinrich-Heine University, 40225 Düsseldorf, Germany
[c] Institute of Developmental Genetics, Cluster of Excellence on Plant Sciences, Heinrich-Heine-University, 40225 Düsseldorf, Germany
[d] Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824
[e] Department of Plant Systems Biology, VIB, B-9052 Gent, Belgium
[f] Department of Plant Biotechnology and Bioinformatics, Ghent University, B-9052 Gent, Belgium

C_4 photosynthesis outperforms the ancestral C_3 state in a wide range of natural and agro-ecosystems by affording higher water-use and nitrogen-use efficiencies. It therefore represents a prime target for engineering novel, high-yielding crops by introducing the trait into C_3 backgrounds. However, the genetic architecture of C_4 photosynthesis remains largely unknown. To define the divergence in gene expression modules between C_3 and C_4 photosynthesis during leaf ontogeny, we generated comprehensive transcriptome atlases of two Cleomaceae species, *Gynandropsis gynandra* (C_4) and *Tarenaya hassleriana* (C_3), by RNA sequencing. Overall, the gene expression profiles appear remarkably similar between the C_3 and C_4 species. We found that known C_4 genes were recruited to photosynthesis from different expression domains in C_3, including typical housekeeping gene expression patterns in various tissues as well as individual heterotrophic tissues. Furthermore, we identified a structure-related module recruited from the C_3 root. Comparison of gene expression patterns with anatomy during leaf ontogeny provided insight into genetic features of Kranz anatomy. Altered expression of developmental factors and cell cycle genes is associated with a higher degree of endoreduplication in enlarged C_4 bundle sheath cells. A delay in mesophyll differentiation apparent both in the leaf anatomy and the transcriptome allows for extended vein formation in the C_4 leaf.

INTRODUCTION

C_4 photosynthesis has evolved concurrently and convergently in angiosperms more than 65 times from the ancestral C_3 state (Sage et al., 2011) and provides fitness and yield advantages over C_3 photosynthesis under permissive conditions, such as high temperatures (Hatch, 1987; Sage, 2004). In brief, C_4 photosynthesis represents a biochemical CO_2 pump that supercharges photosynthetic carbon assimilation through the Calvin-Benson-Bassham cycle (CBBC) by increasing the concentration of CO_2 at the site of its assimilation by the enzyme Rubisco (Andrews and Lorimer, 1987; Furbank and Hatch, 1987). Rubisco is a bifunctional enzyme that catalyzes both the productive carboxylation and the futile oxygenation of ribulose 1,5-bisphosphate. The oxygenation reaction produces a toxic byproduct, 2-phosphoglycolic acid (Anderson, 1971), which is removed by an energy-intensive metabolic repair process called photorespiration. By concentrating CO_2 through the C_4 cycle, the oxygenation of ribulose 1,5-bisphosphate and thereby photorespiration is massively reduced. However, the C_4 cycle requires input of energy to drive the CO_2 pump. Photorespiration increases with temperature and above ~23°C, the energy requirements of metabolic repair become higher than the energy cost of the C_4 cycle (Ehleringer and Björkman, 1978; Ehleringer et al., 1991). Hence, operating C_4 photosynthesis is beneficial at high leaf temperatures, whereas C_3 photosynthesis prevails in cool climates (Ehleringer et al., 1991; Zhu et al., 2008).

With a few exceptions, C_4 photosynthesis requires specialized Kranz anatomy (Haberlandt, 1896), in which two distinct cell types share the photosynthetic labor, namely, mesophyll cells (MCs) and bundle sheath cells (BSCs). MCs surround the BSCs in a wreath-like manner and both cell types form concentric rings around the veins. This leads to a stereotypic vein-BSC-MC-MC-BSC-vein pattern (Brown, 1975). MCs serve as carbon pumps that take in CO_2 from the leaf intercellular air space, convert it into a C_4 carbon compound, and load it into the BSCs. Here, CO_2 is released from the C_4 compound and assimilated into biomass by the CBBC, and the remaining C_3-compound is returned to the MC to be loaded again with CO_2. The carbon

[1] These authors contributed equally to this work.
[2] Address correspondence to andreas.weber@uni-duesseldorf.de.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Andreas P.M. Weber (andreas.weber@uni-duesseldorf.de).
[C] Some figures in this article are displayed in color online but in black and white in the print edition.
[W] Online version contains Web-only data.
[OPEN] Articles can be viewed online without a subscription.
www.plantcell.org/cgi/doi/10.1105/tpc.114.123752

 © 2014 American Society of Plant Biologists. All rights reserved.

pump runs at a higher rate than the CBBC (overcycling), which leads to an increased concentration of CO_2 in the BSCs. Our understanding of the different elements required for C_4 photosynthesis varies, with many components of the metabolic cycle known, while their interplay and regulation remain mostly enigmatic, and very little is known about their anatomical control (Sage and Zhu, 2011).

C_4 photosynthesis can be considered a complex trait, since it requires changes to the expression levels of hundreds or perhaps thousands of genes (Bräutigam et al., 2011, 2014; Gowik et al., 2011). While complex traits are typically dissected by measuring the quantitative variation across a polymorphic population, this approach is not promising for C_4 photosynthesis, due to lack of known plasticity in "C_4-ness" (Sage and McKown, 2006). Historical crosses between C_3 and C_4 plants (Chapman and Osmond, 1974) are no longer available and would have to be reconstructed before they can be analyzed with molecular tools.

Alternatively, closely related C_3 and C_4 species provide a platform for studying C_4 photosynthesis. In the Cleomaceae and Asteraceae, comparative transcriptomic analyses have identified more than 1000 genes differentially expressed between closely related C_3 and C_4 species (Bräutigam et al., 2011; Gowik et al., 2011). These studies, however, compared the end points of leaf development, i.e., fully matured photosynthetic leaves. Therefore, they do not provide insight into the dynamics of gene expression during leaf ontogeny, which is important for understanding the establishment of C_4 leaf anatomy. Systems analyses of maize (*Zea mays*) leaf gradients have provided a glimpse into developmental gene expression modules (Li et al., 2010; Pick et al., 2011; Wang et al., 2013); however, maize lacks a close C_3 relative and has simple parallel venation making any generalizations to dicot leaf development difficult.

Tarenaya hassleriana, previously known as *Cleome hassleriana* (Iltis and Cochrane, 2007; Iltis et al., 2011), which is a C_3 plant, and *Gynandropsis gynandra* (previously known as *Cleome gynandra*), which is a derived C_4 plant, represent an ideal pair for a comparative analysis of the complex trait of C_4 photosynthesis (Bräutigam et al., 2011). Both species belong to the family of Cleomaceae, are closely related to each other and to the well-annotated C_3 plant model species *Arabidopsis thaliana* (Brown et al., 2005; Marshall et al., 2007; Inda et al., 2008), and both Cleome sister lineages share many traits (Iltis et al., 2011). In addition, the genome of *T. hassleriana* has been recently sequenced and serves as a reference for expression profiling via RNA sequencing (Cheng et al., 2013).

In this study, we take advantage of the phylogenetic proximity between *G. gynandra* and *T. hassleriana* to compare the dynamic changes in gene expression during leaf development (Inda et al., 2008). We generated a transcriptome atlas for each species, consisting of three biological replicates of six different stages of leaf development, three different stages of each seed and seedling development, reproductive organs (carpels, stamen, petals, and sepals), stems, and roots. In parallel, we performed microscopy analysis of the leaf anatomy. Finally, we measured leaf cell ploidy levels by flow cytometry and measurements of nuclear size in different leaf cell types by confocal laser scanning microscopy.

RESULTS

Selection of Tissues Featured in the Comparative Atlases

For high-resolution characterization of photosynthetic development between a dicotyledonous C_3 and C_4 species, a leaf developmental gradient was defined. Stage 0 was the youngest sampled leaf, 2 mm in length, and not yet emerged from the apex. The stage 0 leaves are the first to show a discernible palmate shape and contain the first order vein (midrib vein) in both species (Figure 1A; Supplemental Figure 1A). New leaves emerged from the apex every 2 d (plastochron = 2 d) in both species and were numbered sequentially from the aforementioned stage 0 to stage 5 (Figure 1A). The leaves emerge and initiate secondary vein formation at stage 1 (Supplemental Figure 1B) and fully mature by stages 4 and 5 (Supplemental Figures 1E and 1F). The mature leaf of the C_4 species has more minor veins (up to 7°) than that of the C_3 species (up to 6°; Supplemental Figure 1F). The leaf expansion rate is initially indistinguishable and never significantly different between the species (Figure 1B). The sampled leaf gradient covered the development from non-light-exposed sink tissues to fully photosynthetic source tissues.

Complementary to this and to provide a broader comparison between C_3 and C_4 plants, seedlings, minor photosynthetic, and

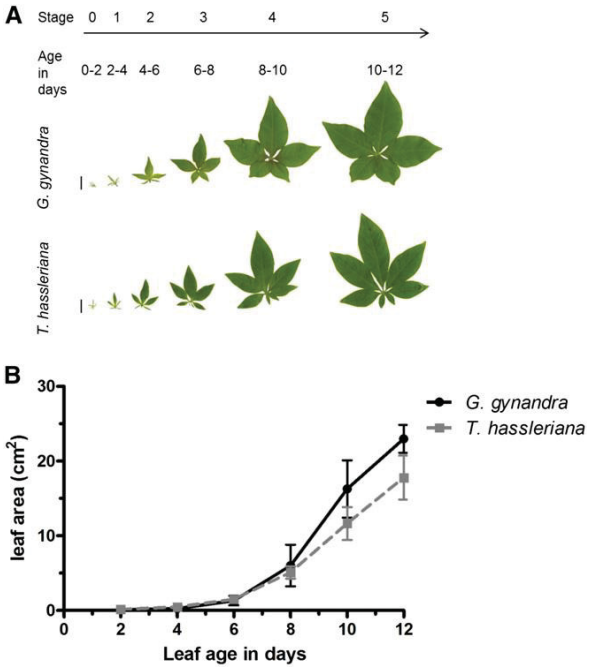


Figure 1. Overview of Leaf Shape and Expansion Rate in *G. gynandra* and *T. hassleriana*.

(A) Image of each leaf category sequenced (bar = 1 cm). Each category is 2 d apart from the other.

(B) Leaf expansion rate of each leaf category in cm^2 over 12 d (n = 5; ±SD)

[See online article for color version of this figure.]

heterotrophic tissues were selected for further characterization. The aerial portion of seedlings (cotyledon and hypocotyl) was sampled 2, 4, and 6 d after germination to cover early cotyledon maturation (Supplemental Figure 2). The full root system and stem tissue were sampled from plants after 6 to 8 weeks of growth before inflorescence emergence (Supplemental Figure 3A); floral organs (petals, carpels, stamen, and sepals) were harvested during flowering of 10- to 14-week-old plants as well as three different stages of seed development (Supplemental Figure 3B). In total, 10 phototrophic and 8 heterotrophic tissues per species were included in the atlases (Table 1).

The C_3 and C_4 Transcriptomes Are of High Quality and Comparable between Species

Cross-species mapping provided a more reliable data set than de novo transcriptome assembly. Between 1.4 and 67 million high-quality reads were generated per replicate (Supplemental Data Set 2). Initially, paired-end reads from each tissue were assembled by VELVET/OASES (Supplemental Table 1). Comparing the resulting contigs to reference data, including the *T. hassleriana* genome (Cheng et al., 2013), revealed several quality issues. These include excessive numbers of contigs mapping to single loci, fused and fragmented contigs, and the absence of C_4 transcripts known to be highly expressed in *G. gynandra* (Supplemental Figures 4A to 4C and Supplemental Data Set 3). As an alternative, we aligned single-end reads from both species to the recently sequenced *T. hassleriana* genome (Cheng et al., 2013). Albeit slightly lower, the mapping efficiency and specificity remained comparable between both species with 60 to 70% of reads mapped for both leaf gradients (Supplemental Data Set 1). To define an upper boundary for any artifacts caused by cross-species mapping, three *T. hassleriana* samples (mature leaf stage 5, stamen, and young seed) were mapped to *Arabidopsis*. The correlation between replicates was equivalent in reads mapped to the cognate genome and across species with an average $r = 0.98$. Furthermore, there was a strong correlation between both mappings, reaching an average Pearson correlation of $r = 0.86$ after collapsing expression data to *Arabidopsis* identifiers to minimize bias from different genome duplication histories (Supplemental Table 2 and Supplemental Figure 5). Cross-species mapping has been successfully used for inter species comparisons before (Bräutigam et al., 2011, 2014; Gowik et al., 2011), and in this study mapping of both species to the *T. hassleriana* genome provided a quality data set with a limited degree of artifacts.

The generated transcriptome atlases were reproducible and comparable between species. To reduce noise, downstream analyses focused on genes expressed above 20 reads per mappable million (RPKM; Supplemental Figure 6), unless otherwise noted. Biological replicates of each tissue clustered closely together and were highly correlated (mean $r = 0.92$, median $r = 0.97$; Figure 2A; Supplemental Figures 7A and 7B and Supplemental Table 3). On average, 4686 and 5308 genes displayed significantly higher expression values in *G. gynandra* and *T. hassleriana*, respectively, with the greatest differences observed in seed and stem tissue (Supplemental Table 4). In contrast, the transcriptome patterns were highly similar between the sister species (Figure 2A; Supplemental Figure 7C). Principle component analysis (PCA) showed that the first component separated the species and accounted for only 15% of the total variation (Supplemental Figure 8A).

Table 1. Sequencing and Mapping Stats for Each Averaged Tissue Sample in *T. hassleriana* and *G. gynandra*

		T. hassleriana			*G. gynandra*		
		Total No. of Reads in Three Replicates	No. of Genes Expressed > 1 RPKM	No. of Genes Expressed > 1000 RPKM	Total No. of Reads in Three Replicates	No. of Genes Expressed > 1 RPKM	No. of Genes Expressed > 1000 RPKM
Leaf gradient	0	58,874,878	23,238	64	75,895,556	22,357	104
	1	59,389,701	23,134	74	66,822,298	22,021	133
	2	63,590,283	23,104	81	55,247,053	22,143	129
	3	90,654,684	23,004	90	75,944,275	21,854	144
	4	36,572,303	22,844	106	69,951,930	21,734	119
	5	102,018,867	22,905	106	69,639,670	21,039	119
Floral organs	Sepal	103,721,357	23,656	74	77,430,418	23,145	83
	Petal	21,754,853	21,379	86	10,872,686	21,322	77
	Stamen	57,929,412	22,642	140	55,748,506	22,489	133
	Carpel	28,021,839	23,910	67	4,929,824	23,577	76
	Stem	30,932,633	23,292	75	59,516,389	22,508	98
	Root	88,911,824	24,255	68	86,879,963	23,430	89
Seedling	2 DAG	90,777,012	23,306	120	89,262,140	21,960	130
	4 DAG	89,517,055	23,041	116	112,658,149	22,036	130
	6 DAG	71,271,739	22,877	138	64,470,699	21,910	136
Seed maturation	1	52,229,844	23,708	118	32,763,383	22,991	118
	2	31,872,067	22,969	145	29,958,720	22,262	148
	3	53,271,349	21,737	138	56,453,325	20,082	152

Reads were normalized as RPKM ($n = 3$). DAG, days after germination.

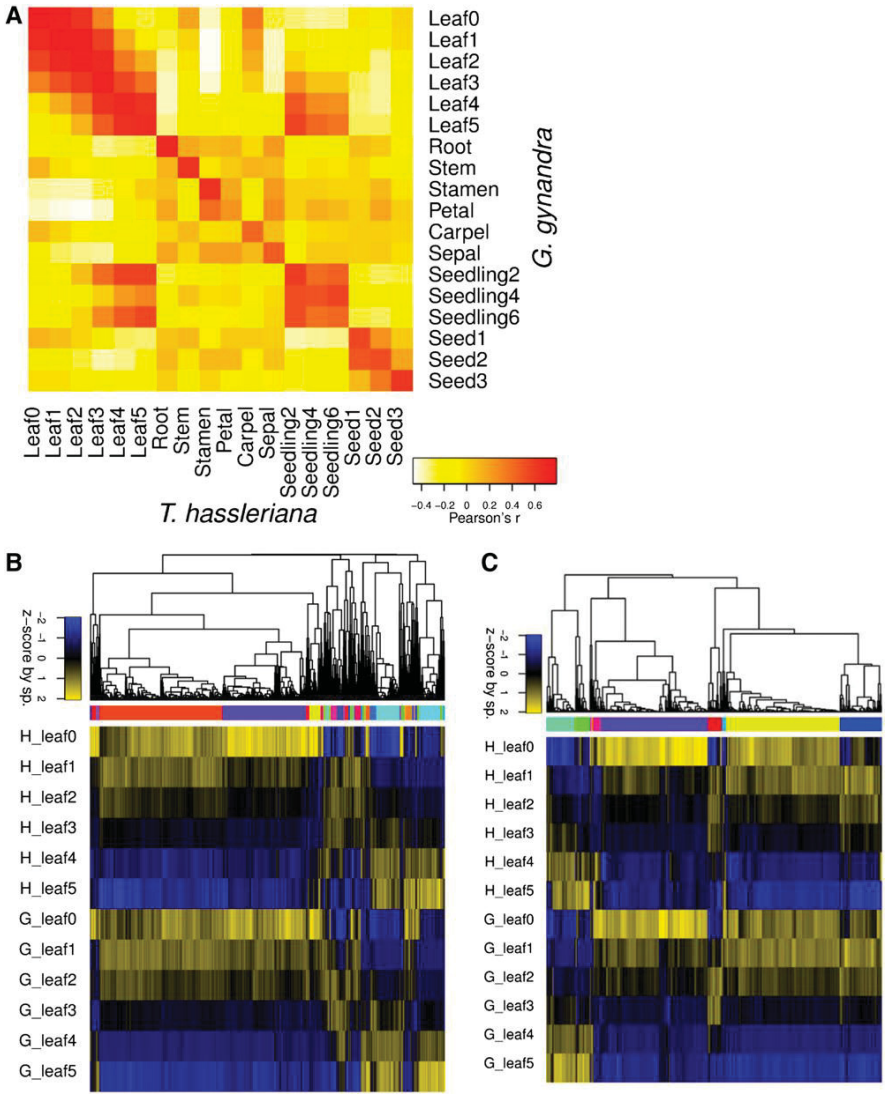


Figure 2. Comparative Tissue Dynamics and Gene Expression Pattern between *G. gynandra* and *T. hassleriana*.

(A) Pearson's correlation heat map of the expression of tissue-specific signature genes (RPKM) of all leaf gradient sample averages (n = 3) per species. Yellow, low expression; red, high expression. G, *G. gynandra*; H, *T. hassleriana*.

(B) Pearson's correlation hierarchical cluster of all leaf gradient sample averages as Z-scores. Blue is the lowest expression and yellow the highest expression.

(C) Expression patterns of transcriptional regulators in both species within the leaf gradient. Pearson's correlation hierarchical cluster of all sample averages as Z-scores. Blue is the lowest expression and yellow the highest expression.

Gene expression patterns and dynamics are conserved between species. The number of genes expressed above 20 RPKM varied by tissue from 6900 to 12,000, with the fewest in the mature leaf and most in the stem and youngest leaf in both species (Table 1; Supplemental Data Set 2). Hierarchical clustering revealed major modules with increasing and decreasing expression along the leaf gradient (Figure 2B), a large overlap of peak expression between seedlings and mature tissue, and distinct gene sets for the other sampled tissues (Supplemental Figure 9A). In leaves, the genes with decreasing expression split into two primary clusters, of which the smaller cluster maintained higher expression longer in the C_4 than the C_3 species (Figure 2B). Clustering of the tissues with 10,000 bootstrap replications confirmed the visual similarity of mature leaves and seedlings and showed further major branches consisting of (1) carpel, stem, and root; (2) a seed gradient and remaining floral

organs; and (3) young leaves (Supplemental Figure 9A). Limiting the clustering to transcription factors (TFs) showed equivalent results (Supplemental Figure 9B; Figure 2C), except that in leaves, a higher proportion of the TFs with decreasing expression maintained expression longer in the C_4 species. Notably, this delay impacted the clustering of the tissues and older C_4 leaves tended to cluster with younger C_3 leaves by TF expression (Supplemental Figures 9A and 9B). The delay was further reflected in a PCA of the leaf gradient where stage 0 and 1 show much less separation in *G. gynandra* than in *T. hassleriana* (Supplemental Figure 8B).

The functional categories with dominant expression showed distinct patterns across the tissues and high conservation between the species. As in the hierarchical clustering, the species showed similar profiles when examining the number of signature genes (expressed over 1000 RPKM; Figure 3) or the total RPKM (Supplemental Figure 9) in each functional category. As expected, in mature leaves and seedlings, transcriptional activity is dominated by photosynthesis, which is almost entirely lacking from roots, seeds, stamens, and petals (Figure 3; Supplemental Figure 9). Younger leaf tissues of the C_3 species show higher expression of genes in the photosynthetic category, displayed as signature genes (Figure 3) or as cumulative RPKM per category (Supplemental Figure 9). In all floral tissues, roots, and stems, transcriptional activity is comparatively balanced between categories. In seeds, a major portion of the total expression is allocated to a few, extremely highly expressed lipid transfer protein type seed storage proteins (Supplemental Figure 9). The differences between the two species lie in the details, especially within the developmental leaf gradient. In young *G. gynandra* leaves, more signature genes encode DNA and protein-associated MapMan terms than in *T. hassleriana* (Figure 3). A close examination of secondary MapMan categories shows that specifically histone proteins (34 genes with $P < 0.05$ in stage 1, enriched with Fisher's exact test $P = 2.6 \cdot 10^{-13}$) and protein synthesis (222 genes with $P < 0.05$ in stage 1, enriched with Fisher's exact test $P = 1.8 \cdot 10^{-17}$) are upregulated in *G. gynandra* and that these categories have a larger dynamic range in *G. gynandra* than *T. hassleriana* (Supplemental Figure 10).

In summary, transcriptomic analysis indicates the tissues are well paired and comparable between species and although there are differences in expression level, there is conservation of expression patterns between species. Within the leaf gradient, there is a subset of genes that shows a delay in the onset of expression changes in *G. gynandra*.

The Comparative Transcriptome Atlases Revealed Diverse Recruitment Patterns from the C_3 Plant *T. hassleriana* to C_4 Photosynthesis

The expression patterns of the core C_4 cycle genes were compared in *G. gynandra* and *T. hassleriana* to gain insight into the evolutionary recruitment of C_4 cycle genes to photosynthesis. During convergent evolution of C_4 photosynthesis, these genes

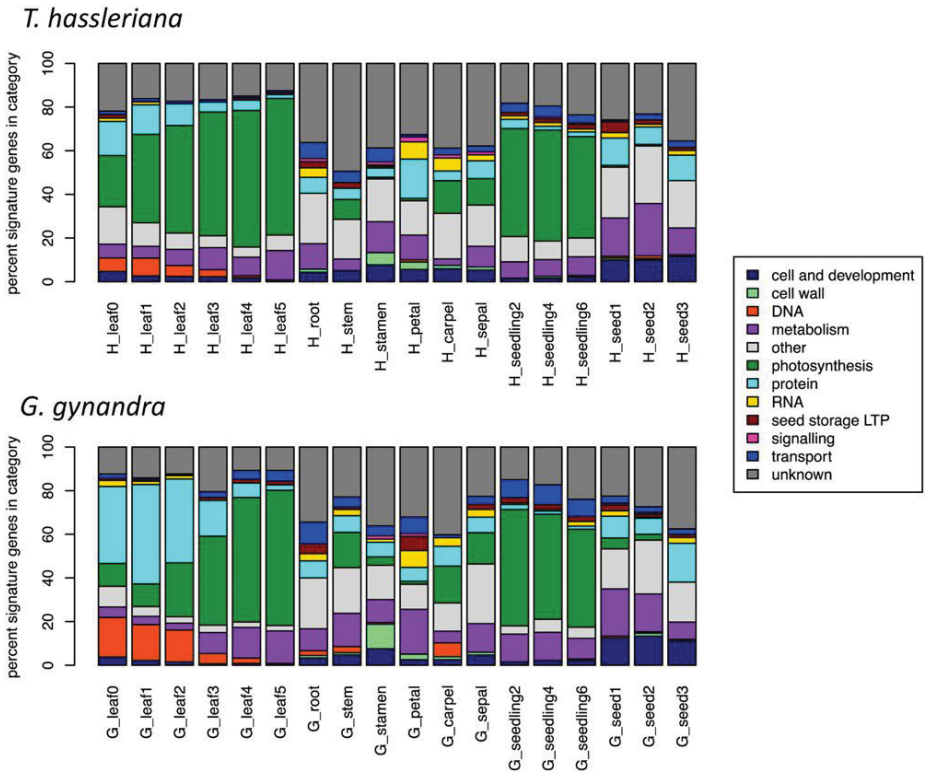


Figure 3. Distribution of Signature Genes in Each Tissue in *G. gynandra* and *T. hassleriana*.

Percentage of signature genes expressed over 1000 RPKM falling in each basal MapMan category for every averaged tissue.

were recruited from ancestral C_3 genes (Sage, 2004; Edwards et al., 2010; Sage et al., 2011). To contextualize the change in expression of the C_4 cycle genes, the between species Euclidean (absolute) and Pearson (pattern) distances were calculated and compared from the leaf developmental gradients (Figure 4A). All known C_4 cycle genes showed a large Euclidean distance (844 to 9156 RPKM), while they split between a correlated and an inversely correlated pattern. In addition to the known C_4 genes, histones, lipid transfer proteins, protein synthesis, and DNA synthesis are functional categories found among genes with greater than 844 RPKM differences in absolute expression (Supplemental Data Set 6).

To identify ancestral C_3 expression domains from which C_4 genes were recruited, the expression of the core C_4 cycle genes was compared between species. In *G. gynandra*, all core C_4 cycle genes increase in expression along the leaf gradient and are high in seedlings (Figures 4C and 4D; Supplemental Figures 12A to 12F); this pattern matches that of other photosynthetic genes (Figure 4B). For each C_4 cycle gene, the *T. hassleriana* sequence to which most *G. gynandra* reads mapped was taken as the most likely closest putative ortholog (Supplemental Figures 13 and 14). The putative orthologs of core C_4 genes are expressed at comparatively low levels in C_3 (Supplemental Figures 13 and 14). Activity measurements of the core C_4 cycle enzymes match the observed gene expression profiles (Supplemental Figure 15). In contrast to leaves and seedlings, the remaining tissues show a variety of expression patterns of C_4 cycle genes in both species (Figures 4C to 4E; Supplemental Figures 12A to 12G). Of the C_4 cycle genes, *NAD-MALIC ENZYME* (*NAD-ME*) and the *SODIUM: HYDROGEN ANTIPORTER* (*NHD*) show a fairly constitutive expression pattern in C_3, while the others have a small number of tissues where the expression peaks (Figure 4C; Supplemental Figure 12A). The expression of *PYRUVATE PHOSPHATE DIKINASE* (*PPDK*), the *PHOSPHO*ENOL*PYRUVATE TRANSLOCATOR* (*PPT*), and *DICARBOXYLATE CARRIER* (*DIC*) peaks in floral organs (Supplemental Figures 12B and 12C; Figure 4D); the expression of *ASPARTATE AMINO TRANSFERASE* (*AspAT*) and *ALANINE AMINOTRANSFERASE* (*AlaAT*) peaks in seed (Figure 4E; Supplemental Figure 12D); and the expression of the pyruvate transporter *BILE ACID:SODIUM SYMPORTER FAMILY PROTEIN2* (*BASS2*) peaks in the young leaf (Supplemental Figure 12E). Albeit erroneous identification of the closest C_3 ortholog in some cases (e.g., *BASS2* and *PHOSPHO*ENOL*PYRUVATE CARBOXYLASE* [*PEPC*]) impedes identification of the ancestral C_3 expression domain (Supplemental Figures 12 and 13), the majority of known C_4 cycle genes were recruited to a photosynthetic expression pattern from a variety of expression domains (Figure 4B).

To assess the possibility of small modular recruitment from other tissues to the C_4 leaf, we searched for evidence of an expression shift between the C_3 root and the C_4 leaf. This shift is expected, if the bundle sheath tissue is partially derived from the regulatory networks of root endodermis, as proposed previously (Slewinski, 2013). Expression pattern filters were used to identify 37 genes that were expressed primarily in the C_3 root and the C_4 leaf (C_3 leaf/root < 0.3; C_4/C3 leaf > 1; C_4 leaf4-5/root > 0.5; C_4 leaf5 > 30 RPKM; leaf5/root enrichment 6-fold greater in C_4), significantly more than in a randomized data set (P value < 10^{-29}; Supplemental Table 5). This set of genes showed a very similar expression pattern to photosynthetic genes along the C_4 leaf gradient (Figure 5A).

The functions encoded by the genes that were apparently recruited to the leaf from a root expression domain were consistent with structural modifications and C_4 photosynthesis. In *Arabidopsis*, 29 of the corresponding homologs are heterogeneously expressed across different root tissues with their highest expression in either the endodermis or cortex, analogous to bundle sheath and mesophyll cells, respectively (Slewinski, 2013). Three functional groups could be identified in the cluster. The first is related to tissue structure, i.e., cell wall modification and plasmodesmata, the second to metabolic flux and redox balance, and the third to signaling (Figure 5B). Among these genes are two C_4 cycle genes, namely, *DIC1*, and a carbonic anhydrase. The group contains three TFs, one of which is involved in auxin response stimulation. Coexpression network analysis of the *Arabidopsis* homologs (ATTED-II) shows 11 genes from the cluster occur in a shared regulatory network. In summary, a set of genes related to cell wall, metabolic/redox flux, and signaling was recruited from the C_3 root to the C_4 leaf, many of which are coexpressed in *Arabidopsis* and found in leaf tissues analogous to BSC and MC.

Changes in the Leaf Transcriptomes Reveal Differences in Cellular Architecture and Leaf Development in the C_4 Species

Altered expression of cell cycle genes and enlarged BSC nuclei in *G. gynandra* suggest the occurrence of endoreduplication within this cell type. During early leaf development, *G. gynandra* leaf samples clustered together with younger samples in *T. hassleriana* (Supplemental Figures 8A and 8B), indicating a delay in leaf maturation. We hypothesized this delay in *G. gynandra* leaf maturation is manifested through alterations of cell cycle gene expression during leaf development. Hierarchical clustering of absolute expression values showed that the majority of known core cell cycle genes (Vandepoele et al., 2002; Beemster et al., 2005) have comparable expression patterns between both species (Supplemental Figure 16 and Supplemental Data Set 7). However, two distinct groups of genes were identified, which are either upregulated in *G. gynandra* between stage 0 to 2 (group 1: 9 of 18 genes with P value < 0.05) or show a delayed decrease during C_4 leaf development (group 2: 9 of 12 genes with P value < 0.05 between stage 0 and 3; Supplemental Figure 16 and Supplemental Data Set 7). Interestingly, *GT-2-LIKE1* (*GTL1*), a key cell cycle regulator, was not correlated between *G. gynandra* and *T. hassleriana* during leaf development. *GTL1* is upregulated in later stages of leaf development in *T. hassleriana* but not in *G. gynandra* (P value < 0.001 in stage 5; Supplemental Figure 16 and Supplemental Data Set 7).

As GTL1 has been demonstrated to operate as an inhibitor of endoreduplication and ploidy-dependent cell growth (Breuer et al., 2009, 2012), we examined whether nuclei were enlarged in any *G. gynandra* leaf tissues. First, both leaf developmental gradients were subjected to flow cytometry. Polyploidy (DNA content > 2C) was observed in both species, but clearly enriched in C_4 compared with C_3, especially in the more mature leaves (5% versus 1% ≥ 8C, 16% versus 4% ≥ 4C; Figure 6A).

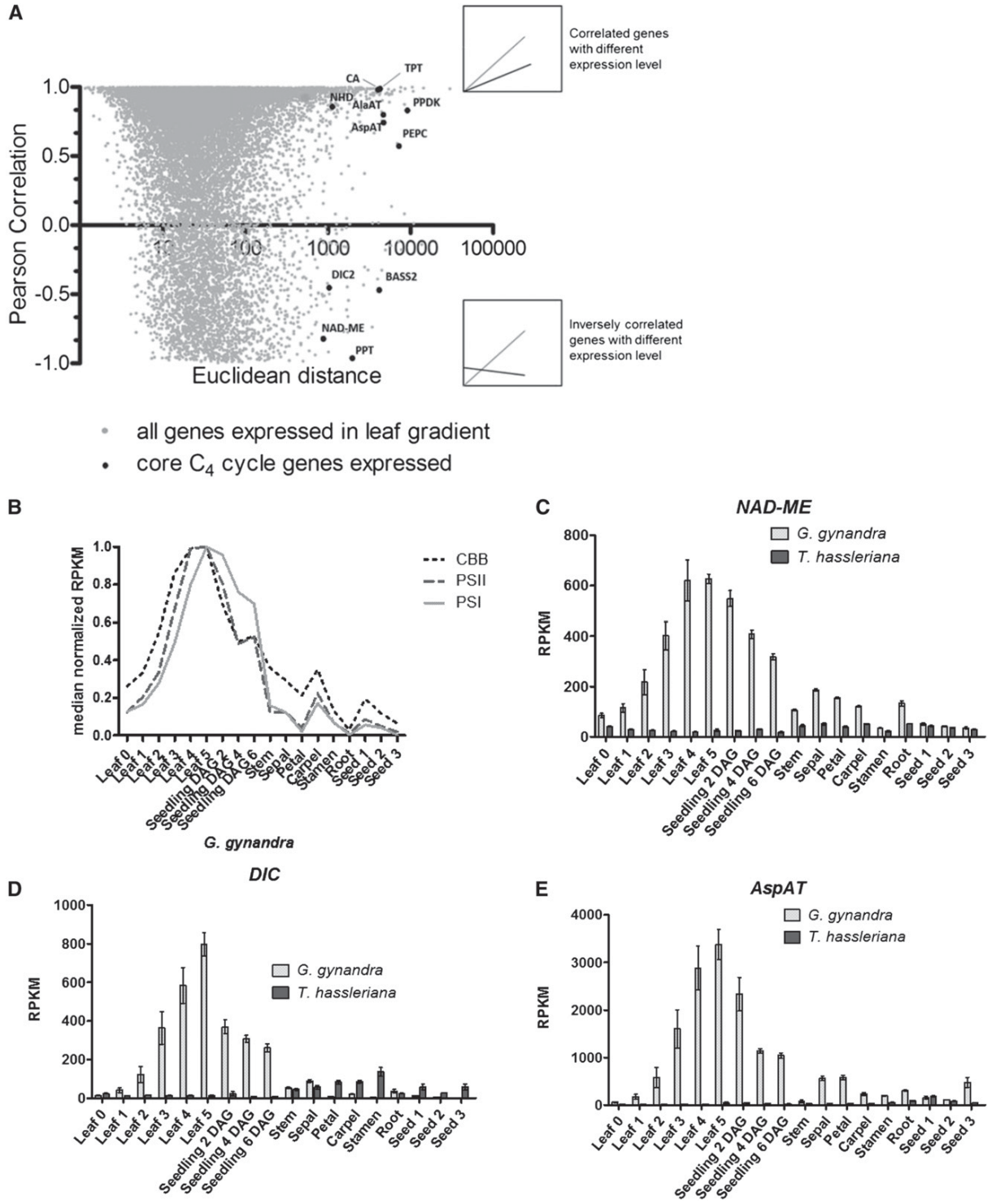


Figure 4. Comparison of Gene Expression Dynamics within the Leaf Gradient of Both Species.

(A) Euclidean distance versus Pearson's correlation of average RPKM (n = 3) of genes expressed (>20 RPKM) in both leaf developmental gradients. Comparison of gene expression by similarity of expression pattern and expression level in *T. hassleriana* and *G. gynandra*. Relevant highly expressed C_4

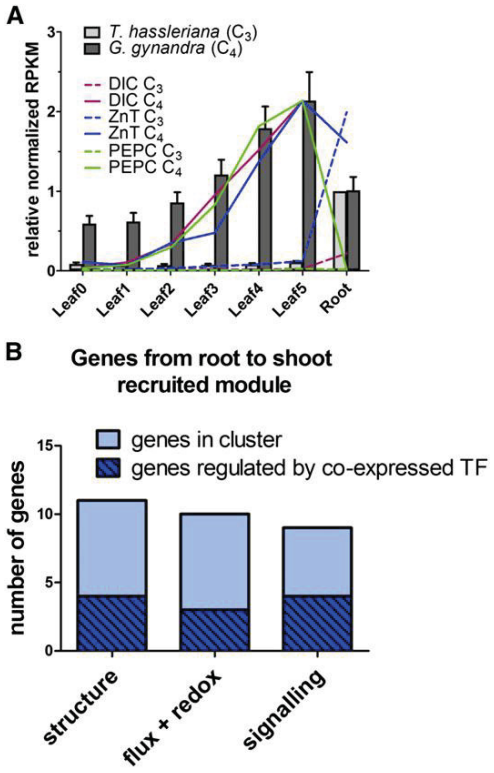


Figure 5. Recruitment of Genes from the Root to Leaf Expression Domain in the C_4 Plant *G. gynandra*.

(A) Relative average RPKM normalized to expression in *G. gynandra* leaf 5 (gray bars). Bars represent the arithmetic means of all 37 genes; lines show expression patterns of a reference C_4 cycle gene (*PEPC*) and of two genes found in the shifted module.
(B) Genes in the module displayed as functional groups. Light blue: absolute number of genes in the group. Dark blue overlay: portion of genes controlled by a transcription factor of the module.

In the *G. gynandra* C_4 leaf, the BSC nuclei were 2.9-fold larger than those in the MC ($P < 0.001$; Figures 6B and 6C). In contrast, the C_3 *T. hassleriana* nuclei of both cell types were similar sizes with a size ratio of 1.0 (Figures 6B and 6C). The proportion of BSC in the leaf was estimated from transversal sections as 15% in *G. gynandra* and 6% *in T. hassleriana* (Figures 7A to 7L). This number fits with the subpopulation of cells with higher ploidy observed in *G. gynandra* in the mature leaf. In summary, the extended expression of a subgroup of cell cycle genes and downregulation of *GTL1* correlate with higher ploidy levels in the *G. gynandra* mature leaf based on BSC nuclei area and flow cytometry measurements.

The C_4 Species Shows Delayed Differentiation of Mesophyll Tissue, Coinciding with Increased Vein Formation

The transcriptional delay in a large subset of *G. gynandra* genes (Figures 2B, 2C, and 3) reflects a later differentiation of the C_4 leaf. The delayed pattern of this large subset of genes indicated that there might be a delay in the differentiation of leaf internal anatomy, although leaf growth rates and shape are similar between species (Figure 1A). Thus, the leaves were examined microscopically. Since dicotyledonous leaves differentiate in a wave from tip toward petiole (Andriankaja et al., 2012), leaves were cross-sectioned at the midpoint (50% leaf length) for comparison. The cross sections revealed that in C_4 leaves, cell differentiation was delayed in the transition from undifferentiated ground tissue toward fully established palisade parenchyma (Figures 7A to 7L). Both species start undifferentiated at leaf stage 0 with only the primary vein distinctly visible in cleared leaves (Figures 7A and 7G; Supplemental Figure 1A). In stage 1, the C_3 leaf starts to differentiate its palisade parenchyma, while the C_4 leaf shows dividing undifferentiated cells (Figures 7B and 7H). Mesophyll differentiation has finished by stage 2 in the C_3 leaf (Figure 7I), but not until stage 4 in the C_4 leaf (Figure 7D). Classical mature C_4 leaf architecture appears in stage 4 in *G. gynandra* (Figure 7E). C_4 leaves ultimately develop more veins and open veinlets leading to Kranz anatomy (Supplemental Figure 1). Leaf mesophyll tissue of the C_3 species differentiates faster and develops fewer veins than the C_4 species.

The expression of genes related to vein development was consistent with greater venation in the C_4 leaf but failed to explain the larger delay in expression patterns and mesophyll differentiation in the C_4 leaf. Hierarchical clustering indicated that most known leaf and vasculature developmental factors (reviewed in Ohashi-Ito and Fukuda, 2010) showed similar expression patterns in the two species (Supplemental Figure 17 and Supplemental Table 6). However, two clusters with distinct expression patterns were detected. In the C_4 species, seven genes were upregulated (P value < 0.05), including vasculature facilitators *PIN-FORMED* (*PIN1*), *HOMEOBOX GENE8* (*HB8*), and *XYLOGEN PROTEIN1* (*XYP1*) (Motose et al., 2004; Scarpella et al., 2006; Donner et al., 2009), while five genes were downregulated (P value < 0.05), among those the negative regulators *KANADI1* and *2*, as well as *HOMEOBOX GENE15* (Supplemental Figure 17 and Supplemental Table 6; Ilegems et al., 2010).

To further elucidate the magnitude and nature of the delayed expression changes on the transcriptional level, the leaf gradient data were clustered with the *K*-means algorithm (Supplemental

Figure 4. (continued).

cycle genes are marked in plot. Above inset shows an example of two highly correlated genes by expression trend and strength. Lower inset shows an example of two genes inversely correlated with different expression level.
(B) Expression pattern across the atlas of averaged relative expression of transcripts encoding for photosystem I (PSI), photosystem II (PSII), and soluble enzymes of the Calvin-Benson-Bassham (CBB) cycle in *G. gynandra.*
(C) to **(E)** Average expression pattern of highest abundant ortholog of C_4 cycle genes (*NAD-ME*, *DIC*, and *AspAT*) in photo- and heterotrophic tissues in *G. gynandra* (light gray) and *T. hassleriana* (dark gray); ±SE, $n = 3$.

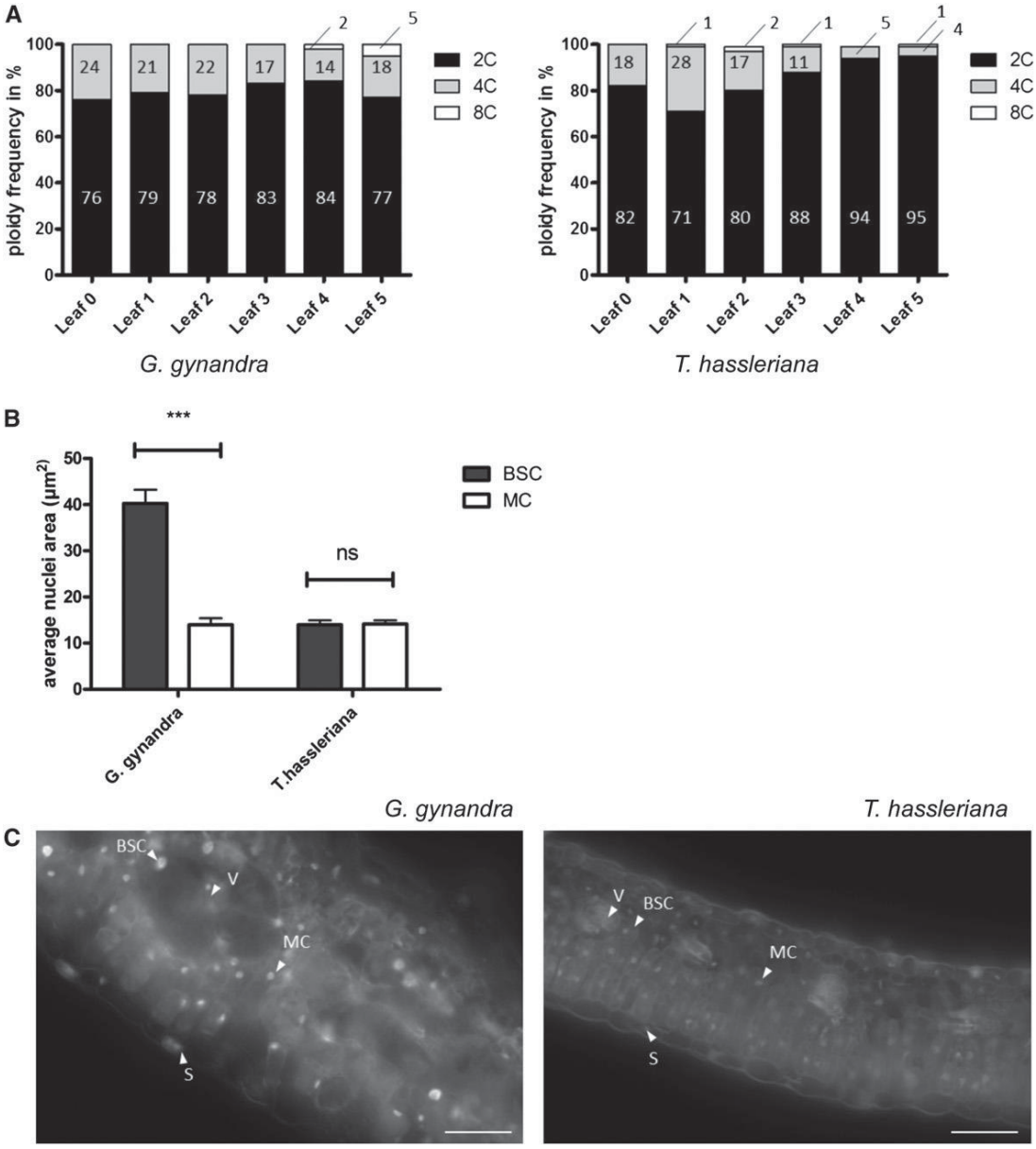


Figure 6. Distribution of Ploidy Levels during Leaf Development and Nuclei Area of BSC and MC between *G. gynandra* and *T. hassleriana*.

(A) Ploidy distribution of developing leaf (category 0 till 5) in percentage in *G. gynandra* and *T. hassleriana*. Measurements performed in $n = 3$ (except G0 = 1 replicate). For each replicate, at least 2000 nuclei were measured by flow cytometry.

(B) Quantification of BSC and MC nuclei area in cross sections ($n = 3 \pm$ SE) of mature *G. gynandra* and *T. hassleriana* leaves (stage 5). Area of nuclei in μm^2 with at least 150 nuclei analyzed per cell type per species per replicate. Asterisks indicate statistically significant differences between BSC and MC (***P value < 0.001); n.s., not significant.

(C) Fluorescence microscopy images of propidium iodide-stained leaf cross sections (stage 5) of *T. hassleriana* (left) and *G. gynandra* (right). Arrowheads point to nuclei of the indicated cell type. V, vein; S, stomata. Bar = 50 µm.

Figures 17A and 17B and Supplemental Data Set 9). Of 16 clusters, six were divergent (1 to 3, 8, 9, and 15; 1270 genes). The remaining clusters were similar; however, four showed a transcriptional delay (4, 5, 13, and 16; 3361 genes), while six did not (6, 7, 10 to 12, and 14; 5162 genes). Of all clustered genes, 87% belonged to highly conserved clusters, 34% with a delay and 53% without. Thus, the transcriptional delay cannot be explained by general slower development.

All of the *K*-means clusters were functionally characterized by testing for enrichment in MapMan categories (Supplemental

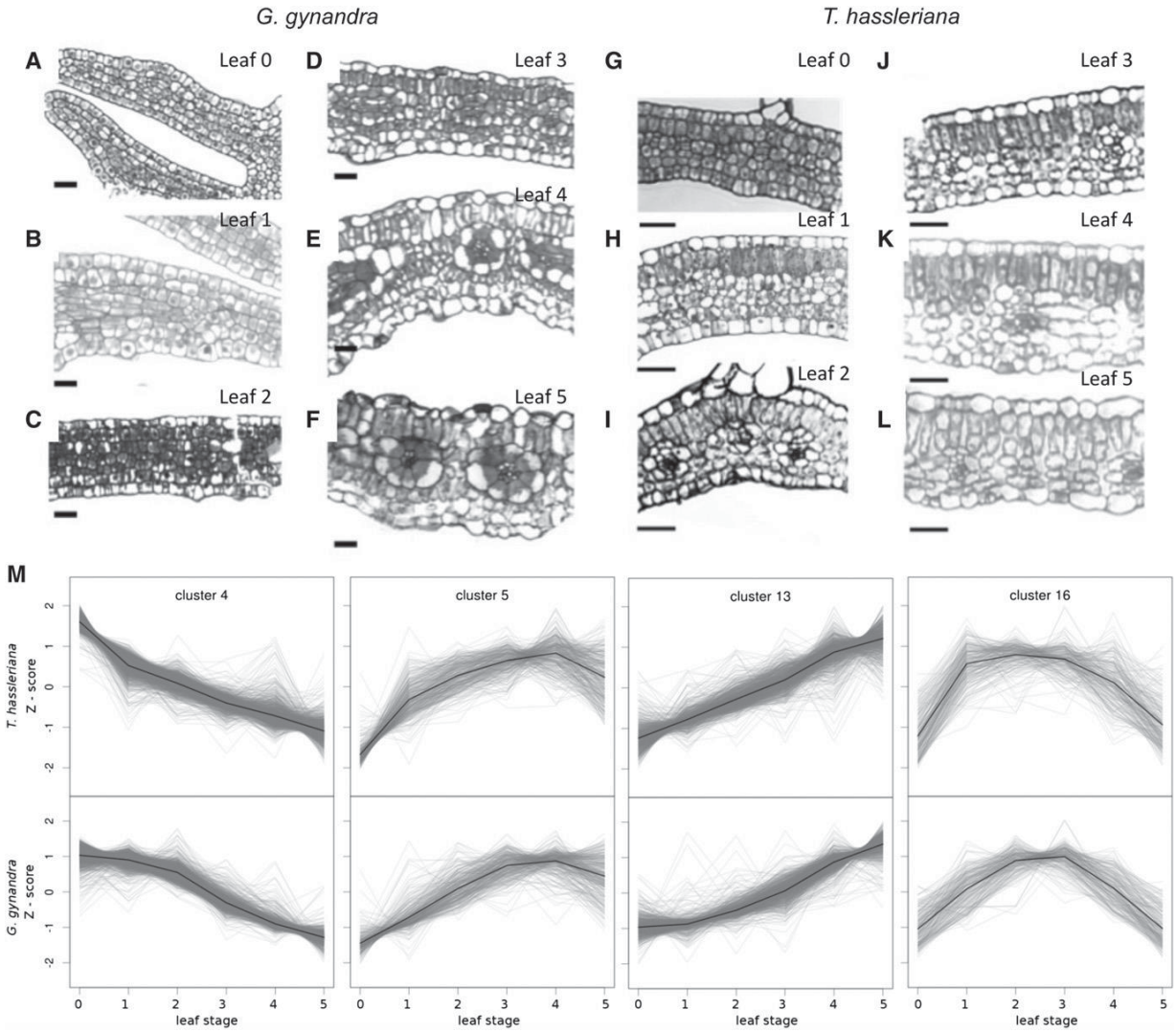


Figure 7. Analysis of Shifted Gene Expression Pattern and Leaf Anatomy during Leaf Ontogeny.

(A) to **(L)** Leaf anatomy development along the gradient in *G. gynandra* and *T. hassleriana* depicted by cross sections stained with toluidine blue. Bar = 20 μm.

(M) Selected clusters from *K*-means clustering of gene expression shown as *Z*-scores, which show a phase shift between *G. gynandra* and *T. hassleriana* during leaf development.

Data Set 10). The visually "shifted" patterns were: later onset of increase in clusters 13 and 5 (1058 and 395 genes, respectively), delayed decrease in cluster 4 (1644 genes), and a later peak in cluster 16 (264 genes; Figure 7M). The "late decrease" cluster 4 is enriched in genes related to mitochondrial electron transfer, *CONSTITUTIVE PHOTOMORPHOGENESIS9* (*COP9*) signalosome, and protein degradation by the proteasome (Figure 7M; Supplemental Data Set 10). The "late onset" cluster 13 is enriched in all major photosynthetic categories: N-metabolism, and chlorophyll, isoprenoid, and tetrapyrrole biosynthesis (P value < 0.05; Supplemental Figures 17C and 17D and Supplemental Data Sets 9 and 10). The smaller "late onset" cluster 5 is enriched in the categories protein synthesis, tetrapyrrole synthesis, carotenoids, and peroxiredoxin. Cluster 16 peaks earlier in *T. hassleriana* than *G. gynandra* and is enriched in lipid metabolism (e.g., *ACYL CARRIER PROTEIN4*, *CHLOROPLASTIC ACETYLCOA CARBOXYLASE1*, *3-KETOACYL-ACYL CARRIER PROTEIN SYNTHASE1*, and *3-KETOACYL-ACYL CARRIER PROTEIN SYNTHASE III*) and plastid division genes, such as the *FILAMENTATION TEMPERATURE-SENSITIVE* genes *FtsZ2*, *FtsH*, and *FtsZ*, as well as *ACCUMULATION AND REPLICATION OF CHLOROPLASTS11* (Figure 7M; Supplemental Data Sets 9 and 10).

The core of the phase-shifted clusters, defined as genes with Pearson's correlation coefficient of $r > 0.99$ to the cluster center, contained candidate regulators for the observed delayed patterns. The core of cluster 13 contained 17 TFs and genes involved in chloroplast maintenance (Supplemental Data Set 11). The core of cluster 4 contained 30 transcriptional regulators, including *PROPORZ1* (*PRZ1*), and eight other chromatin-remodeling genes. Nineteen cell cycle genes were found in the core of cluster 4 (Supplemental Figures 19A and 19B), including *CELL DIVISION CYCLE20* (*CDC20*), *CDC27*, and *CELL CYCLE SWITCH PROTEIN52* (*CCS52*), which are key components of cell cycle progression from M-phase to S-phase (Pérez-Pérez et al., 2008; Mathieu-Rivet et al., 2010b).

Our data were quantitatively compared with data from *Arabidopsis* leaf development to test if the observed phase shift related to a switch from proliferation to differentiation (Andriankaja et al., 2012). This study identified genes that were significantly up- or downregulated during the shift from proliferation to expansion (Andriankaja et al., 2012). Putative orthologs of these genes were clustered by the *K*-means algorithm (without prior expression filtering), producing seven clusters for the upregulated genes (containing 483 genes in total) and five clusters for the downregulated genes (1112 genes in total; Supplemental Figure 20). The trend was well conserved across species, with 75% of the upregulated and 96% of the downregulated genes falling into clusters with a matching trend. The genes showed a higher proportion of delay in *G. gynandra* than in the total data set, with 60 and 68% falling in delayed up- and downregulated clusters, respectively (Supplemental Figure 20).

In summary, about a third of all gene expression patterns show a delay in the *G. gynandra* leaf (Figure 7M; Supplemental Figure 18). Delayed genes include major markers of leaf maturity such as the upregulation of photosynthetic gene expression and downregulation of mitochondrial electron transport (Supplemental Figures 19C and 19D and Supplemental Data Set 10). This delay was more common in putative orthologs of genes differentially regulated during the shift from cell proliferation to expansion (Supplemental Figure 19; Andriankaja et al., 2012). The slow maturation can be seen on the anatomical level as a delayed differentiation that coincides with increased vein formation in the C_4 species (Figures 7A to 7L).

DISCUSSION

Comparative Transcriptome Atlases Provide a Powerful Tool for Understanding C_4 Photosynthesis

Two transcriptome atlases were generated to allow the analysis of gene recruitment to photosynthesis and to detect differences related to C_4 leaf anatomy. Two Cleomaceae species were chosen for this study due to their phylogenetic proximity to the model species *Arabidopsis* (Marshall et al., 2007). The sampled leaf tissues covered development from sink tissue to fully mature source tissue (Figures 1 and 3), and all higher order vein development (Supplemental Figure 1). Since C_4 genes are recruited from genes already present in C_3 ancestors, where they carry out housekeeping functions (Sage, 2004; Besnard et al., 2009; Christin and Besnard, 2009; Christin et al., 2009), seed, stem, floral, and root tissues were included in the atlases in addition to leaves and seedlings.

The high similarity in expression pattern between the species maximizes our ability to detect differences related to C_4 photosynthesis. While PCA analysis showed that the first principle component separated the data set by species, this accounted for only 15% of the variation (Supplemental Figure 8A). Excluding floral organs and stem, all tissues correlated with $r > 0.7$ between species (Supplemental Figure 7C and Supplemental Table 3). Hierarchical and *K*-means clustering showed the vast majority of genes had a similar pattern between species, and tissue types clustered closely with the same tissue in the other species. Specific groups of highly expressed genes exclusively expressed in one tissue type, such as root, stamen, and petal, are shared between *G. gynandra* and *T. hassleriana*, suggesting that these genes might represent drivers for the respective tissue identity (Supplemental Figure 9). A subset of genes showed a consistent adjustment to their expression pattern, namely, a delay in the leaf gradient of *G. gynandra* relative to *T. hassleriana* (Figure 7M). Thus, organ identity is highly conserved between *G. gynandra* and *T. hassleriana*, but the rate at which organ identity, especially the leaf, is established can differ.

Expression Patterns of C_3 Putative Orthologs Support Small-Scale or Modular Recruitment to Photosynthesis, Implying That a General C_4 Master Regulator Is Unlikely

Ancestral expression patterns can be compared with assess whether a master regulator could have facilitated recruitment of genes to C_4 photosynthesis. The patterns of gene expression in *T. hassleriana* provide a good proxy for the ancestral C_3 expression pattern due to its phylogenetic proximity to *G. gynandra* (Inda et al., 2008; Cheng et al., 2013). Genes active in the C_4 cycle were recruited from previously existing metabolism (Matsuoka, 1995; Chollet et al., 1996; Streatfield et al., 1999; Wheeler et al., 2005; Tronconi et al., 2010). Expression patterns in *T. hassleriana* reflect known metabolism and expression; for instance, *PPDK* is expressed in seeds, stamens, and petals (Supplemental Figure 12B), which is similar to the expression domain reported by Chastain et al. (2011). Furthermore, *PPT* is highly expressed in stamens and during seed development (Supplemental Figure 12C; Knappe et al., 2003a, 2003b), since it is required for fatty acid production (Hay and Schwender, 2011).

The C_3 putative orthologs of C_4 cycle genes show a variety of expression patterns within the atlas, providing strong evidence they could not have been recruited by a single master regulator. All C_4 cycle genes are expressed to a low degree in *T. hassleriana*, either constitutively or in defined tissues such as stamens, seeds, or young leaves (Figures 4C to 4E). Expression of *NHD*, *AlaAT*, *AspAT*, and *PPDK* increased along the leaf gradient in both C_3 and C_4 species, but in C_3, the expression was highest in tissues other than the leaf (Figure 4E; Supplemental Figures 12A, 12B, and 12D). In contrast, *DIC*, *BASS2*, *NAD-ME*, and *PPT* are expressed in inverse patterns between C_3 and C_4 along the leaf gradient (Figures 4C and 4D; Supplemental Figures 12C and 12E), and *PEPC* is expressed only in mature leaves in the C_3 species (Supplemental Figure 12F). Except for *DIC* and *PPDK*, the expression level of the C_4 cycle genes was higher in *G. gynandra* across all tissues (Figure 4; Supplemental

Figures 12 to 14). Thus, most of the C_4 cycle genes may still maintain their ancestral functions in addition to the acquired C_4 function. The correct ortholog in C_3 may not have been conclusively determined by cross species read mapping in all cases reported here. However, the main conclusion—that C_4 cycle genes are recruited from a variety of C_3 expression patterns—holds regardless of which putative C_3 paralog is selected (Supplemental Figures 13 and 14).

A set of genes shifted from a root to leaf expression domain during C_4 evolution provides an example of small-scale modular recruitment. The proposed analogy between root endodermis and bundle sheath and between root cortex and mesophyll (Slewinski, 2013) has been linked to cooption of the *SCARECROW* (*SCR*) and *SHORTROOT* (*SHR*) regulatory networks into developing leaves (Slewinski et al., 2012; Wang et al., 2013). A set of 37 genes consistent with such a recruitment module was identified. For this gene set, the C_3 species *T. hassleriana* (Figure 5; Supplemental Table 5) and *Arabidopsis* (Brady and Provart, 2009) showed conserved root expression, while the C_4 species showed an expression pattern similar to photosynthesis. Much of the root to leaf gene set was coregulated in *Arabidopsis*, and it contained TFs, including *ETHYLENE RESPONSE FACTOR1* (Mantiri et al., 2008), as well as an AUX/IAA regulator (Pérez-Pérez et al., 2010) and *VND-INTERACTING2* (Yamaguchi et al., 2010). Functionally, the majority of the gene set is involved in processes related to cell wall synthesis and modification. The set contains the cell wall-plasma membrane linker protein (Stein et al., 2011) and the xyloglucan endotransglycosylase *TOUCH4* (Xu et al., 1995), the tonoplast intrinsic protein involved in cell elongation (Beebo et al., 2009), and a plasmodesmata-located protein (Bayer et al., 2008). The observed coregulation and structural functions support an underlying structural relationship between the root tissues endodermis and cortex, and the leaf tissues bundle sheath and mesophyll.

It is still unresolved whether expression level recruitment of genes to the C_4 cycle was facilitated by the action of one or a few master switches controlling C_4 cycle gene expression and/or by changes to promoter sequences of C_4 genes (Westhoff and Gowik, 2010). The diverse transcriptional patterns of the core C_4 cycle genes in *T. hassleriana* provide strong evidence that they were not recruited as a single transcriptional module facilitated by one or a few master regulators. However, the identified root to leaf module indicates that small-scale corecruitment occurs, and this may help bring about the 3 to 4% overall transcriptional changes occurring during C_4 evolution (Bräutigam et al., 2011, Gowik et al., 2011). The similarities in expression pattern between photosynthetic genes and C_4 cycle genes are evident (Figure 4B), and light-dependent induction of C_4 genes has been reported (Christin et al., 2013), leading us to hypothesize that C_4 cycle genes may use the same light-induced regulatory circuits employed for the photosynthetic genes, possibly through acquisition of *cis*-regulatory elements or modification of chromatin structure, as has been shown for the *PEPC* gene promoter in maize (Tolley et al., 2012).

Cell Size in *G. gynandra* Coincides with Nuclei Size and Ploidy

In addition to the biochemical C_4 cycle genes, transcriptional changes related to cell and tissue architecture are required for C_4 leaf development (Westhoff and Gowik, 2010). The comparative atlases were contextualized with anatomical data to better understand BSC size.

G. gynandra has generally larger cells (Figures 7A to 7L), which might be attributed to a larger genome. After divergence from *T. hassleriana*, the *G. gynandra* lineage has undergone a putative whole-genome duplication (Inda et al., 2008). Cell size has been tied to genome ploidy status previously (Sugimoto-Shirasu and Roberts, 2003; Lee et al., 2009b; Chevalier et al., 2011). A relationship between ploidy and cell size could explain the generally larger cells in *G. gynandra* leaves (Figures 7A to 7L) or relate to the upregulation of DNA and histone-associated genes in developing leaves (Figure 3; Supplemental Figures 10 and 11).

Changes in the expression of key cell cycle genes indicated endoreduplication may be increased in *G. gynandra*, and follow-up nuclear size measurements indeed indicate BSCs have undergone endoreduplication. Enlargement of BSC is a common feature of C_4 plants (Sage, 2004; Christin et al., 2013) including *G. gynandra* (Figures 7D to 7F), but the genetic mechanism is unknown. During leaf development, key cell cycle genes showed changes in expression pattern and expression level between *G. gynandra* and *T. hassleriana* (Supplemental Figure 16). *CDC20* and *CCS52A*, which are closely linked with cell cycle M-to-S-phase progression or endocycle onset (Lammens et al., 2008; Larson-Rabin et al., 2009; Kasili et al., 2010; Mathieu-Rivet et al., 2010a), exhibit prolonged expression during C_4 leaf development, whereas the expression of the master endoreduplication regulator *GTL1* (Breuer et al., 2009, 2012; Caro et al., 2012) is suppressed in the older leaf stages (Supplemental Figure 16). Although a comparison of the more distantly related species *Arabidopsis* and *G. gynandra* discounted endoreduplication as a factor in bundle sheath cell size (Aubry et al., 2013), the BSC and MC nuclei area measurements of mature *G. gynandra* and *T. hassleriana* leaves revealed that the BSC nuclei are 2.9-fold enlarged compared with MC nuclei in *G. gynandra* (Figures 6B and 6C). At the same time, *T. hassleriana* BSC and MC cells do not differ significantly in nuclei size (Figures 6A and 6C).These results are supported by a flow cytometry analysis of both leaf developmental gradients, where the proportion of endoreplicated cells in the mature C_4 leaf (Figures 6A) matches the number of BSCs present in *G. gynandra* (Figures 6A and 7A to F). Interestingly, we also find significant ($P > 0.001$) enlarged BSC nuclei in other C_4 species (e.g., *Flaveria trinervia*, *Megathyrsus maximum*, and maize; Supplemental Figure 22), indicating that larger nuclei size in BSC compared with the MC could be a general phenomenon in C_4 plants conserved across mono- and dicotyledons. Whether endoreplication is the cause of increased cell size in C_4 BSC, as found for trichomes and tomato (*Solanum lycopersicum*) karyoplasm (Traas et al., 1998; Chevalier et al., 2011) or whether endoreplication only occurs to support the high metabolic activity and large size of the BSCs (Sugimoto-Shirasu and Roberts, 2003) remains to be determined.

Late Differentiation of Mesophyll Tissue Allows Denser Venation

General regulators of leaf anatomy and shape (reviewed in Byrne, 2012) are expressed in very similar patterns between the two species (Supplemental Figure 17), reflecting the very similar

palmate five-fingered leaf shape and speed of leaf expansion (Figures 1A and 1B). However, anatomical studies of leaf development show that differentiated palisade parenchyma is already observed at the midpoint of stage 1 leaves in *T. hassleriana* (Figure 7H) but can only be detected in the middle of the leaf in stages 3 and 4 in *G. gynandra* (Figures 7D to 7F). Hierarchical clustering of transcriptome data indicates a similarity between younger *T. hassleriana* and older *G. gynandra* tissues (Supplemental Figure 9), which we attribute to a delay in *G. gynandra* leaf expression changes observed in the hierarchical clusters (Figures 2B and 2C) and observed for *K*-means clustering involving about a third of clustered genes (Figure 7M; Supplemental Figure 18). Analysis of the delayed clusters for significant enrichment of functional categories indicated that the metabolic shift from sink to source tissue was delayed (Figures 3 and 7M; Supplemental Figure 18 and Supplemental Data Set 10). Furthermore, the "delayed decrease" cluster 4 was enriched in *COP9* signalosome and marker genes of the still developing heterotrophic leaf.

Cell cycle and cell differentiation regulators show a delayed expression pattern in *G. gynandra*. The expression of *PRZ1*, which switches development from cell proliferation to differentiation in *Arabidopsis* (Sieberer et al., 2003; Anzola et al., 2010), is prolonged in the C_4 leaf (Figure 7M, cluster 4), as is the expression of chromatin remodeling factor *GRF1-INTERACTING FACTOR3* implicated in the control of cell proliferation upstream of cell cycle regulation (Lee et al., 2009a). Plastid division genes peak around leaf stage 1 in *T. hassleriana* and leaf stage 2 in *G. gynandra* (Figure 7M, cluster 16). It has recently been shown that chloroplast development and division precedes photosynthetic maturity in *Arabidopsis* leaves and retrograde signaling from the chloroplasts affects cell cycle exit from proliferation (Andriankaja et al., 2012). Quantitative comparison of differentially regulated genes during the shift from cell proliferation to cell expansion found in *Arabidopsis* (Supplemental Figure 20; Andriankaja et al., 2012) to the expression patterns of the putatively orthologous genes along leaf developmental gradients in Cleome, reveals a strong conservation of expression pattern between *Arabidopsis* and Cleome during development. A higher proportion of delay of *G. gynandra* genes is observed in this gene set. This supports the idea that the transcriptional delay is directly linked to the anatomical delay in differentiation observed in *G. gynandra* (Supplemental Figure 19).

The delay in cell differentiation allows for increased vein formation in the C_4 leaf. Mesophyll differentiation has already been shown to limit minor vein formation in *Arabidopsis* (Scarpella et al., 2004; Kang et al., 2007). *G. gynandra* and *T. hassleriana* have altered vein densities, which result from more minor vein orders in *G. gynandra* (Supplemental Figure 1), similar to results for the dicot *Flaveria* species (McKown and Dengler, 2009). Given that differentiation of photosynthetic mesophyll cells limits minor vein formation in *Arabidopsis* (Scarpella et al., 2004; Kang et al., 2007) and that mesophyll differentiation is delayed in the C_4 species compared with the C_3 species (Figure 7), dense venation may indeed be achieved by delaying mesophyll differentiation.

Genes related to vascular patterning are expressed in a manner consistent with higher venation in the C_4 leaf. The high expression of vascular pattern genes such as *PIN1*, *HB8*, *ARF3*, and *XYP1* in the C_4 leaf (Supplemental Figure 17) is similar to that observed for Kranz patterned leaves in maize (Wang et al., 2013). However, these genes may be a consequence, rather than a cause, of higher venation, especially since some of these markers are only expressed after pre-procambial or procambial identity is introduced (Ohashi-Ito and Fukuda, 2010). Once procambial fate is established, cellular differentiation of vein tissues proceeds through positional cues and localized signaling, possibly via the SCR/SHR pathway (Langdale and Nelson, 1991; Nelson and Langdale, 1992; Nelson and Dengler, 1997; Griffiths et al., 2013; Wang et al., 2013; Lundquist et al., 2014). Interestingly, in accordance with the delay in leaf differentiation in *G. gynandra*, we could monitor a delay in higher expression for *SHR* peaking around leaf stage 1 to 3 (Supplemental Figure 21A). *SCR* transcript abundance is clearly divided in both *G. gynandra* and *T. hassleriana* between two homologs, one of which is more abundant in the C_4 leaf and the other in the C_3 leaf (Supplemental Figure 21B). *SCR* expression in *G. gynandra* follows the *SHR* pattern with a delayed upregulation. This is in accordance with earlier studies conducted in maize, where *SHR* transcript highly accumulates in the BSC to activate *SCR* expression (reviewed in Slewinski et al., 2012)

The identification of mesophyll differentiation as the proximate cause for fewer minor vein orders in *T. hassleriana* raises the question of how mesophyll differentiation is controlled. In both C_4 and C_3 species, vascular patterning precedes photosynthetic tissue differentiation (Sud and Dengler, 2000; Scarpella et al., 2004; McKown and Dengler, 2010). Light is one of the most important environmental cues that regulate leaf development, including its cellular differentiation and onset of photosynthesis (Tobin and Silverthorne, 1985; Nelson and Langdale, 1992; Fankhauser and Chory, 1997). The *COP9* signalosome, which plays a central role in repression of photomorphogenesis and G2/M cell cycle progression (Chamovitz et al., 1996; Dohmann et al., 2008), showed a delayed decrease in *G. gynandra* compared with *T. hassleriana* (Supplemental Figure 19B). The delay and earlier vein formation termination induced by excess light in *Arabidopsis* (Scarpella et al., 2004) suggest that light perception and its signal transduction may be differentially regulated in species with denser venation patterns.

Conclusions

In this study, we report a detailed comparison of the transcriptomes and the leaf development of two Cleomaceae species with different modes of photosynthetic carbon assimilation, i.e., C_3 and C_4 photosynthesis. The gene expression patterns are quite similar between both species, which facilitates the identification of differences related to C_4 photosynthesis. We could link two key features of Kranz anatomy to developmental processes through integration of expression and anatomical data. First, we show that the larger size of the bundle sheath cells in the C_4 species is associated with a higher ploidy in these cells, which might be controlled by delayed repression of the endocycle via the transcription factor *GTL1*. Second, a prominent difference between C_3 and C_4 leaf development is the delayed differentiation of the leaf cells in C_4, which is associated with a delayed onset of photosynthetic gene expression, chloroplast proliferation and development, and altered expression of a few

distinct cell cycle genes. Delayed mesophyll differentiation allows for increased initiation of vascular tissue and thus contributes to the higher vein density in C_4. We hypothesize that delayed onset of mesophyll and chloroplast differentiation is a consequence of the prolonged expression of the *COP9* signalosome and, hence, a delayed derepression of photomorphogenesis.

METHODS

Plant Material and Growth Conditions

Gynandropsis gynandra and *Tarenaya hassleriana* plants for transcriptome profiling by Illumina Sequencing were grown in standard potting mix in a greenhouse between April and August 2011. Internal transcribed spacer sequences of *G. gynandra* and *T. hassleriana* were analyzed and plant identity confirmed according to Inda et al. (2008). Leaves were harvested from 4- to 6-week-old plants, prior to inflorescence initiation. All samples were harvested during midday. Flowers, stamens, sepals, and carpels were harvested after induction of flowering. Green tissues from seedlings were harvested 2, 4, and 6 d after germination. Root material was harvested from plants grown in vermiculite for 6 weeks and supplemented with Hoagland solution. Leaf material for the ontogeny analysis was selected by the order of leaf emergence from the apex in leaf stages from 0 to 5. Up to 40 plants were pooled for each biological replicate.

Leaf Expansion Rate

Leaves from stage 0 to 5 were analyzed in five biological replicates for each *G. gynandra* and *T. hassleriana*. Leaves were scanned on a flat bed scanner (V700 Photo; Epson), and the area was analyzed with free image analysis software ImageJ.

Leaf Cross Sections for Anatomical Studies

Leaves from stage 0 to 5 were analyzed in biological triplicates. Leaf material (2 × 2 mm) was cut next to the major first order vein at 50% of the whole leaf length. Leaf material was fixed in 4% paraformaldehyde solution overnight at 4°C, transferred to 0.1% glutaraldehyde in phosphate buffer, and vacuum infiltrated three times for 5 min. The leaf material was then dehydrated with an ascending ethanol series (70, 80, 90, and 96%) with a 1-h incubation in each solution. Samples were incubated twice in 100% ethanol and twice in 100% acetone, each for 20 min, and infiltrated with an acetone:araldite (1:1) mixture overnight at 4°C. After acetone evaporation, fresh araldite was added to the leaf samples until samples were covered and incubated for 3 to 4 h. Samples were transferred to fresh araldite in molds and polymerized at 65°C for 48 h. Cross sections were stained with toluidine blue for 15 s and washed with H_2O_{dest}. Cross sections were imaged with bright-field settings using an Eclipse Ti-U microscope (Nikon).

Flow Cytometry

Three biological replicate samples were chopped with a razor blade in 200 μL of Cystain UV Precise P Nuclei extraction buffer followed by the addition of 800 μL of staining buffer (buffers from Partec). The chopped leaves in buffer were filtered through a 50-μm mesh. The distribution of the nuclear DNA content was analyzed using a CytoFlow ML flow cytometer and FLOMAX software (Partec) as described (Zhiponova et al., 2013).

Measurement of Nuclei from Mature Leaves

Fresh mature leaves (leaf stage 5, three biological replicates) of *G. gynandra* and *T. hassleriana* were cut transversally, fixed in 1× PBS buffer (1% Tween 20 and 3% glutaraldehyde) overnight at room temperature, and stained with propidium iodide solution directly on the microscopic slide. Cross sections were imaged by fluorescence microscopy using an Axio Imager M2M fluorescence microscope (Zeiss) with an HE DS-Red Filter. Images were processed with ZEN10 software (Zeiss), and the nuclear area of at least 200 nuclei per cell type per species was measured with ImageJ.

RNA Extraction, Library Construction, and Sequencing

Plant material was extracted using the Plant RNeasy extraction kit (Qiagen). RNA was treated on-column (Qiagen) and in solution with RNA-free DNase (New England Biolabs). RNA integrity, sequencing library quality, and fragment size were checked on a 2100 Bioanalyzer (Agilent). Libraries were prepared using the TruSeq RNA Sample Prep Kit v2 (Illumina), and library quantification was performed with a Qubit 2.0 (Invitrogen). Single-end sequenced samples were multiplexed with six libraries per lane with ~20 million reads per library. For paired-end sequencing, RNA of all photosynthetic and nonphotosynthetic samples was pooled equally for each species and prepared as one library per species. Paired end libraries were run on one lane with ~175 million clean reads for *T. hassleriana* and 220 million clean reads for *G. gynandra*. All libraries were sequenced on the HISEQ2000 Illumina platform. Libraries were sequenced in the single-end or paired-end mode with length ranging from 80 to 100 nucleotides. The paired-end library of *G. gynandra* had an average fragment size of 304 bp; *T. hassleriana* had an average fragment size of 301 bp.

Gene Expression Profiling

Reads were checked for quality with FASTQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/), subsequently cleaned and filtered for quality scores greater than 20 and read length greater than 50 nucleotides using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit). Expression abundances were determined by mapping the single-end read libraries (each replicate for each tissue) independently against *T. hassleriana* representative coding sequences (Cheng et al., 2013) using BLAT V35 (Kent, 2002) in protein space and counting the best mapping hit based on e-value for each read uniquely. Default BLAT parameters were used for mapping both species. Expression was normalized to reads per kilobase *T. hassleriana* coding sequence per million mappable reads (RPKM). *T. hassleriana* coding sequences were annotated using BLASTX searches (cutoff $1e^{-10}$) against the TAIR10 proteome database. The best BLAST hit per read was filtered by the highest bit score. A threshold of 20 RPKM per coding sequence in at least one species present in at least one tissue was chosen to discriminate background transcription (Supplemental Figure 14). Differential expression between *T. hassleriana* and *G. gynandra* was determined by EdgeR (Robinson et al., 2010) in R (R Development Core Team, 2009). A significance threshold of 0.05 was applied after the P value was adjusted with false discovery rate via Bonferroni-Holms correction (Holm, 1979).

Data Analysis

Data analysis was performed with the R statistical package (R Development Core Team, 2009) unless stated otherwise. For Pearson's correlation and PCA analysis, *Z*-scores were calculated by gene across both species. For all other analyses, *Z*-scores were calculated by gene within each species, to focus on comparing expression patterns. For *K*-means and hierarchical clustering, genes were filtered to those with more than 20 RPKM in at least one of the samples used in each species. To determine the number of centers for *K*-means clustering, the sum of SE within clusters was plotted against cluster number and compared with randomized data (Supplemental Figures 18B, 20C, and 20D). A total of 16 centers was chosen, and *K*-means clustering was performed 10,000 times and the best solution, as defined by the minimum sum of SE of genes in the cluster, was taken for downstream analyses (Peeples, 2011). Multiscale bootstrap resampling of the hierarchical clustering was

performed for samples with 10,000 repetitions using the pvclust R package (Suzuki and Shimodaira, 2006).

Stage enrichment was tested for all *K*-means clusters and for tissue "signature genes" with expression of over 1000 RPKM in each tissue using TAIR10 MapMan categories (from http://mapman.gabipd.org) for the best *Arabidopsis thaliana* homolog. Categories with more than five members in the filtered (*K*-means) or complete (signature genes) data set were tested for enrichment by Fisher's exact test, and P values were adjusted to false discovery rates via Benjamini-Yekutieli correction, which is tolerant of dependencies (Yekutieli and Benjamini, 1999).

Accession Numbers

Sequence data from this article can be found in NCBI GenBank under the following accession numbers: SRP036637 for *G. gynandra* and SRP036837 for *T. hassleriana*.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Venation Patterning during Leaf Development of *G. gynandra* and *T. hassleriana*.

Supplemental Figure 2. *G. gynandra* Cotyledon Anatomy 2, 4, and 6 d after germination (DAG).

Supplemental Figure 3. Images of Tissues Harvested for Atlases in *G. gynandra* and *T. hassleriana*.

Supplemental Figure 4. Quality Assessment of Velvet/OASES Assembled *T. hassleriana* Contigs against Predicted Corresponding CDS from *T. hassleriana* Genome

Supplemental Figure 5. Quality Assessment of the Biological Replicates of *T. hassleriana* Libraries Mapped to *A. thaliana* and Mapping Similarity of *T. hassleriana* Libraries Mapped to *A. thaliana* and to Its Own CDS.

Supplemental Figure 6. Determination of Baseline Gene Expression via a Histogram of Photosystem (PS) I and II Transcript Abundances (RPKM) in the *G. gynandra* Root.

Supplemental Figure 7. Quality Assessment of the Biological Replicates within Each Species and Tissue Similarity between *G. gynandra* and *T. hassleriana*.

Supplemental Figure 8. Principle Component Analysis between *G. gynandra* and *T. hassleriana.*

Supplemental Figure 9. Hierarchical Cluster Analysis with Bootstrapped Samples of *G. gynandra* and *T. hassleriana.*

Supplemental Figure 10. Transcriptional Investment of Each Tissue Compared in Both Species.

Supplemental Figure 11. Transcriptional Investment at Secondary MapMan Category Level of Each Tissue Compared in Both Species.

Supplemental Figure 12. Comparison of Gene Expression Dynamics within the Leaf Gradient of Both Species.

Supplemental Figure 13. Plot of the Expression Pattern (RPKM) of all C_4 Gene Orthologs Expression Pattern in *G. gynandra.*

Supplemental Figure 14. Plot of the Expression Pattern of all C_4 Gene Putative Orthologs Expression Pattern (RPKM) in *T. hassleriana.*

Supplemental Figure 15. Enzyme Activity Measurement of Soluble C_4 Cycle Enzymes.

Supplemental Figure 16. Hierarchical Clustering of Average RPKM with Euclidean Distance of Core Cell Cycle Genes.

Supplemental Figure 17. Hierarchical Clustering with Pearson's Correlation of Leaf Developmental Factors.

Supplemental Figure 18. *K*-Means Clustering of Leaf Gradient Expression Data and Quality Assessment.

Supplemental Figure 19. Z-Score Plots of Enriched MapMan Categories in the Shifted Clusters.

Supplemental Figure 20. *K*-Means Clustering of Genes Differentially Regulated during the Transition from Proliferation to Enlargement.

Supplemental Figure 21. Transcript Abundances of *SCARECROW* and *SHORTROOT* Homologs in *G. gynandra* and *T. hassleriana* Leaf and Root.

Supplemental Figure 22. Nuclei Area and Images of C_4 and C_3 Species.

Supplemental Table 1. Velvet/OASES Assembly Stats from *G. gynandra* and *T. hassleriana* Paired-End Reads.

Supplemental Table 2. Cross-Species Mapping Results.

Supplemental Table 3. Pearson's Correlation between G. *gynandra* and *T. hassleriana* Individual Tissues.

Supplemental Table 4. Number of Significantly Up- or Downregulated Genes in *G. gynandra* Compared with *T. hassleriana* within the Different Tissues.

Supplemental Table 5. List of Genes Present in Root-to-Shoot Recruitment Module.

Supplemental Table 6. List of Clustered General Leaf Developmental and Vasculature Regulating Genes along Both Leaf Gradients.

Supplemental Methods

The following materials have been deposited in the DRYAD repository under accession number http://dx.doi.org/10.5061/dryad.8v0v6.

Supplemental Data Set 1. Annotated Transcriptome Expression Data of Both Atlases in RPKM.

Supplemental Data Set 2. Sequencing and Mapping Statistics for All Single-End Libraries Sequenced.

Supplemental Data Set 3. Quality Assessment of Representative Contigs against Predicted CDS within *T. hassleriana*.

Supplemental Data Set 4. MapMan Categories of Highly Expressed Genes in Each Tissue.

Supplemental Data Set 5. Transcriptional Investment of Each Enriched Basal MapMan Categories in Percentage for Each Tissue.

Supplemental Data Set 6. List of All Genes with Euclidean Distance over 800 RPKM Expressed within Both Leaf Gradients.

Supplemental Data Set 7. List of Core Cell Cycle Genes Selected for Clustering.

Supplemental Data Set 8. Statistical Analysis of Differential Transcript Abundances between *G. gynandra* and *T. hassleriana* for Each Tissue.

Supplemental Data Set 9. Genes Assigned by *K*-Means Clustering to Each Cluster.

Supplemental Data Set 10. MapMan Enrichment Analysis of *K*-Means Clustering.

Supplemental Data Set 11. List of Genes Highly Correlated with Cluster Centers of Shifted Clusters.

ACKNOWLEDGMENTS

Work in our laboratory was supported by grants from the Deutsche Forschungsgemeinschaft (EXC 1028, IRTG 1525, and WE 2231/9-1 to

A.P.M.W.). We thank the HHU Biomedical Research Center (BMFZ) for support with RNA-seq analysis and the MSU High Performance Computing Cluster for support with computational analysis of RNA-seq data. We thank Stefanie Weidtkamp-Peters and the HHU Center for Advanced Imaging for expert advice and support with confocal microscopy and image analysis.

AUTHOR CONTRIBUTIONS

C.K. performed experimental work, analyzed data, and wrote the article. A.K.D. analyzed data and cowrote the article. M.S. assisted in data analysis, identified the root-to-shoot shift, and cowrote the article. J.M., S.S., T.J.W., and E.G.-C. assisted in data analysis. B.B. assisted in design of ploidy experiments. C.R.B assisted in data analysis and experimental design. R.S. assisted in data discussion. L.D.V. assisted in ploidy determination. A.B. analyzed data and wrote the article. A.P.M.W. designed the study and wrote the article.

Received January 30, 2014; revised June 20, 2014; accepted July 6, 2014; published August 8, 2014.

REFERENCES

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

Andrews, T.J., and Lorimer, G.H. (1987). Rubisco: Structure, mechanisms, and prospects for improvement. In The Biochemistry of Plants, Vol. 10, Photosynthesis, M.D. Hatch and N.K. Boardman, eds (San Diego, CA: Academic Press), pp. 131–218.

Andriankaja, M., Dhondt, S., De Bodt, S., Vanhaeren, H., Coppens, F., De Milde, L., Mühlenbock, P., Skirycz, A., Gonzalez, N., Beemster, G.T.S., and Inzé, D. (2012). Exit from proliferation during leaf development in *Arabidopsis thaliana*: a not-so-gradual process. Dev. Cell **22:** 64–78.

Anzola, J.M., Sieberer, T., Ortbauer, M., Butt, H., Korbei, B., Weinhofer, I., Müllner, A.E., and Luschnig, C. (2010). Putative Arabidopsis transcriptional adaptor protein (PROPORZ1) is required to modulate histone acetylation in response to auxin. Proc. Natl. Acad. Sci. USA **107:** 10308–10313.

Aubry, S., Knerová, J., and Hibberd, J.M. (2013). Endoreduplication is not involved in bundle-sheath formation in the C_4 species *Cleome gynandra.* J. Exp. Bot. **65:** 3557–3566.

Bayer, E., Thomas, C., and Maule, A. (2008). Symplastic domains in the Arabidopsis shoot apical meristem correlate with PDLP1 expression patterns. Plant Signal. Behav. **3:** 853–855.

Beebo, A., et al. (2009). Life with and without AtTIP1;1, an Arabidopsis aquaporin preferentially localized in the apposing tonoplasts of adjacent vacuoles. Plant Mol. Biol. **70:** 193–209.

Beemster, G.T.S., De Veylder, L., 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 Physiol. **138:** 734–743.

Besnard, G., Baali-Cherif, D., Bettinelli-Riccardi, S., Parietti, D., and Bouguedoura, N. (2009). Pollen-mediated gene flow in a highly fragmented landscape: consequences for defining a conservation strategy of the relict Laperrine's olive. C. R. Biol. **332:** 662–672.

Brady, S.M., and Provart, N.J. (2009). Web-queryable large-scale data sets for hypothesis generation in plant biology. Plant Cell **21:** 1034–1051.

Bräutigam, A., et al. (2011). An mRNA blueprint for C4 photosynthesis derived from comparative transcriptomics of closely related C3 and C4 species. Plant Physiol. **155:** 142–156.

Bräutigam, A., Schliesky, S., Külahoglu, C., Osborne, C.P., and Weber, A.P.M. (2014). Towards an integrative model of C_4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C_4 species. J. Exp. Bot. **65:** 3579–3593.

Breuer, C., Morohashi, K., Kawamura, A., Takahashi, N., Ishida, T., Umeda, M., Grotewold, E., and Sugimoto, K. (2012). Transcriptional repression of the APC/C activator CCS52A1 promotes active termination of cell growth. EMBO J. **31:** 4488–4501.

Breuer, C., Kawamura, A., Ichikawa, T., Tominaga-Wada, R., Wada, T., Kondou, Y., Muto, S., Matsui, M., and Sugimoto, K. (2009). The trihelix transcription factor GTL1 regulates ploidy-dependent cell growth in the Arabidopsis trichome. Plant Cell **21:** 2307–2322.

Brown, N.J., Parsley, K., and Hibberd, J.M. (2005). The future of C4 research—maize, Flaveria or Cleome? Trends Plant Sci. **10:** 215–221.

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

Byrne, M.E. (2012). Making leaves. Curr. Opin. Plant Biol. **15:** 24–30.

Caro, E., Desvoyes, B., and Gutierrez, C. (2012). GTL1 keeps cell growth and nuclear ploidy under control. EMBO J. **31:** 4483–4485.

Chamovitz, D.A., Wei, N., Osterlund, M.T., von Arnim, A.G., Staub, J.M., Matsui, M., and Deng, X.W. (1996). The COP9 complex, a novel multisubunit nuclear regulator involved in light control of a plant developmental switch. Cell **86:** 115–121.

Chapman, E.A., and Osmond, C.B. (1974). The effect of light on the tricarboxylic acid cycle in green leaves: III. A Comparison between some C(3) and C(4) plants. Plant Physiol. **53:** 893–898.

Chastain, C.J., Failing, C.J., Manandhar, L., Zimmerman, M.A., Lakner, M.M., and Nguyen, T.H.T. (2011). Functional evolution of C(4) pyruvate, orthophosphate dikinase. J. Exp. Bot. **62:** 3083–3091.

Cheng, S., et al. (2013). The Tarenaya hassleriana genome provides insight into reproductive trait and genome evolution of crucifers. Plant Cell **25:** 2813–2830.

Chevalier, C., Nafati, M., Mathieu-Rivet, E., Bourdon, M., Frangne, N., Cheniclet, C., Renaudin, J.-P., Gévaudant, F., and Hernould, M. (2011). Elucidating the functional role of endoreduplication in tomato fruit development. Ann. Bot. (Lond.) **107:** 1159–1169.

Chollet, R., Vidal, J., and O'Leary, M.H. (1996). Phosphoenolpyruvate carboxylase: A ubiquitous, highly regulated enzyme in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **47:** 273–298.

Christin, P.-A., and Besnard, G. (2009). Two independent C4 origins in Aristidoideae (Poaceae) revealed by the recruitment of distinct phosphoenolpyruvate carboxylase genes. Am. J. Bot. **96:** 2234–2239.

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

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

Dohmann, E.M.N., Levesque, M.P., De Veylder, L., Reichardt, I., Jürgens, G., Schmid, M., and Schwechheimer, C. (2008). The Arabidopsis COP9 signalosome is essential for G2 phase progression and genomic stability. Development **135:** 2013–2022.

Donner, T.J., Sherr, I., and Scarpella, E. (2009). Regulation of preprocambial cell state acquisition by auxin signaling in Arabidopsis leaves. Development **136:** 3235–3246.

Edwards, E.J., et al.; C4 Grasses Consortium (2010). The origins of C4 grasslands: integrating evolutionary and ecosystem science. Science **328:** 587–591.

Ehleringer, J.R., and Björkman, O. (1978). A comparison of photosynthetic characteristics of encelia species possessing glabrous and pubescent leaves. Plant Physiol. **62:** 185–190.

Ehleringer, J.R., Sage, R.F., Flanagan, L.B., and Pearcy, R.W. (1991). Climate change and the evolution of C(4) photosynthesis. Trends Ecol. Evol. (Amst.) **6:** 95–99.

Fankhauser, C., and Chory, J. (1997). Light control of plant development. Annu. Rev. Cell Dev. Biol. **13:** 203–229.

Furbank, R.T., and Hatch, M.D. (1987). Mechanism of c(4) photosynthesis: the size and composition of the inorganic carbon pool in bundle sheath cells. Plant Physiol. **85:** 958–964.

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

Griffiths, H., Weller, G., Toy, L.F., and Dennis, R.J. (2013). You're so vein: bundle sheath physiology, phylogeny and evolution in C3 and C4 plants. Plant Cell Environ. **36:** 249–261.

Haberlandt, G. (1896). Physiologische Pflanzenanatomie. (Leipzig, Germany: Verlag von Wilhelm Engelmann).

Hatch, M.D. (1987). C-4 photosynthesis - a unique blend of modified biochemistry, anatomy and ultrastructure. Biochim. Biophys. Acta **895:** 81–106.

Hay, J., and Schwender, J. (2011). Computational analysis of storage synthesis in developing *Brassica napus* L. (oilseed rape) embryos: flux variability analysis in relation to ^{13}C metabolic flux analysis. Plant J. **67:** 513–525.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scand. J. Stat. **6:** 65–70.

Ilegems, M., Douet, V., Meylan-Bettex, M., Uyttewaal, M., Brand, L., Bowman, J.L., and Stieger, P.A. (2010). Interplay of auxin, KANADI and Class III HD-ZIP transcription factors in vascular tissue formation. Development **137:** 975–984.

Iltis, H.H., and Cochrane, T.S. (2007). Studies in the Cleomaceae V: A new genus and ten new combinations for the flora of North America. Novon **17:** 447–451.

Iltis, H.H., Hall, J.C., Cochrane, T.S., and Sytsma, K.J. (2011). Studies in the Cloemaceae I. On the separate recognition of Capparaceae, Cleomaceae, and Brassicaceae. Annals Miss. Bot. Gard. **98:** 28–36.

Inda, L.A., Torrecilla, P., Catalán, P., and Ruiz-Zapata, T. (2008). Phylogeny of Cleome L. and its close relatives Podandrogyne Ducke and Polanisia Raf. (Cleomoideae, Cleomaceae) based on analysis of nuclear ITS sequences and morphology. Plant Sys. Evol. **274:** 111–126.

Kang, J., Mizukami, Y., Wang, H., Fowke, L., and Dengler, N.G. (2007). Modification of cell proliferation patterns alters leaf vein architecture in *Arabidopsis thaliana*. Planta **226:** 1207–1218.

Kasili, R., Walker, J.D., Simmons, L.A., Zhou, J., De Veylder, L., and Larkin, J.C. (2010). SIAMESE cooperates with the CDH1-like protein CCS52A1 to establish endoreplication in *Arabidopsis thaliana* trichomes. Genetics **185:** 257–268.

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

Knappe, S., Flügge, U.I., and Fischer, K. (2003a). Analysis of the plastidic phosphate translocator gene family in Arabidopsis and identification of new phosphate translocator-homologous transporters, classified by their putative substrate-binding site. Plant Physiol. **131:** 1178–1190.

Knappe, S., Löttgert, T., Schneider, A., Voll, L., Flügge, U.I., and Fischer, K. (2003b). Characterization of two functional phosphoenolpyruvate/phosphate translocator (PPT) genes in Arabidopsis—AtPPT1 may be involved in the provision of signals for correct mesophyll development. Plant J. **36:** 411–420.

Lammens, T., Boudolf, V., Kheibarshekan, L., Zalmas, L.P., Gaamouche, T., Maes, S., Vanstraelen, M., Kondorosi, E., La Thangue, N.B., Govaerts, W., Inzé, D., and De Veylder, L. (2008). Atypical E2F activity restrains APC/CCCS52A2 function obligatory for endocycle onset. Proc. Natl. Acad. Sci. USA **105:** 14721–14726.

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

Larson-Rabin, Z., Li, Z., Masson, P.H., and Day, C.D. (2009). FZR2/CCS52A1 expression is a determinant of endoreduplication and cell expansion in Arabidopsis. Plant Physiol. **149:** 874–884.

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

Lee, H.O., Davidson, J.M., and Duronio, R.J. (2009b). Endoreplication: polyploidy with purpose. Genes Dev. **23:** 2461–2477.

Li, P., et al. (2010). The developmental dynamics of the maize leaf transcriptome. Nat. Genet. **42:** 1060–1067.

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

Mantiri, F.R., Kurdyukov, S., Lohar, D.P., Sharopova, N., Saeed, N.A., Wang, X.-D., Vandenbosch, K.A., and Rose, R.J. (2008). The transcription factor MtSERF1 of the ERF subfamily identified by transcriptional profiling is required for somatic embryogenesis induced by auxin plus cytokinin in *Medicago truncatula*. Plant Physiol. **146:** 1622–1636.

Marshall, D.M., Muhaidat, R., Brown, N.J., Liu, Z., Stanley, S., Griffiths, H., Sage, R.F., and Hibberd, J.M. (2007). Cleome, a genus closely related to Arabidopsis, contains species spanning a developmental progression from C(3) to C(4) photosynthesis. Plant J. **51:** 886–896.

Mathieu-Rivet, E., Gévaudant, F., Cheniclet, C., Hernould, M., and Chevalier, C. (2010a). The anaphase promoting complex activator CCS52A, a key factor for fruit growth and endoreduplication in tomato. Plant Signal. Behav. **5:** 985–987.

Mathieu-Rivet, E., Gévaudant, F., Sicard, A., Salar, S., Do, P.T., Mouras, A., Fernie, A.R., Gibon, Y., Rothan, C., Chevalier, C., and Hernould, M. (2010b). Functional analysis of the anaphase promoting complex activator CCS52A highlights the crucial role of endo-reduplication for fruit growth in tomato. Plant J. **62:** 727–741.

Matsuoka, M. (1995). The gene for pyruvate, orthophosphate dikinase in C_4 plants: structure, regulation and evolution. Plant Cell Physiol. **36:** 937–943.

McKown, A.D., and Dengler, N.G. (2009). Shifts in leaf vein density through accelerated vein formation in C4 Flaveria (Asteraceae). Ann. Bot. (Lond.) **104:** 1085–1098.

McKown, A.D., and Dengler, N.G. (2010). Vein patterning and evolution in C-4 plants. Botany **88:** 775–786.

Motose, H., Sugiyama, M., and Fukuda, H. (2004). A proteoglycan mediates inductive interaction during plant vascular development. Nature **429:** 873–878.

Nelson, T., and Langdale, J.A. (1992). Developmental genetics of C-4 photosynthesis. Annu. Rev. Plant Physiol. Plant Mol. Biol. **43:** 25–47.

Nelson, T., and Dengler, N. (1997). Leaf vascular pattern formation. Plant Cell **9:** 1121–1135.

Ohashi-Ito, K., and Fukuda, H. (2010). Transcriptional regulation of vascular cell fates. Curr. Opin. Plant Biol. **13:** 670–676.

Peeples, M.A. (2011). R Script for K-Means Cluster Analysis. http://www.mattpeeples.net/kmeans.html.

Pérez-Pérez, J.M., Candela, H., Robles, P., López-Torrejón, G., del Pozo, J.C., and Micol, J.L. (2010). A role for AUXIN RESISTANT3 in the coordination of leaf growth. Plant Cell Physiol. **51:** 1661–1673.

Pérez-Pérez, J.M., Serralbo, O., Vanstraelen, M., González, C., Criqui, M.C., Genschik, P., Kondorosi, E., and Scheres, B. (2008). Specialization of CDC27 function in the *Arabidopsis thaliana* anaphase-promoting complex (APC/C). Plant J. **53:** 78–89.

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

R Development Core Team (2009). R: A Language and Environment for Statistical Computing. (Vienna, Austria: R Foundation for Statistical Computing).

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

Sage, R.F. (2004). The evolution of C-4 photosynthesis. New Phytol. **161:** 341–370.

Sage, R.F., and McKown, A.D. (2006). Is C4 photosynthesis less phenotypically plastic than C3 photosynthesis? J. Exp. Bot. **57:** 303–317.

Sage, R.F., Christin, P.A., and Edwards, E.J. (2011). The $C_{(4)}$ plant lineages of planet Earth. J. Exp. Bot. **62:** 3155–3169.

Sage, R.F., and Zhu, X.G. (2011). Exploiting the engine of $C_{(4)}$ photosynthesis. J. Exp. Bot. **62:** 2989–3000.

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

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

Sieberer, T., Hauser, M.T., Seifert, G.J., and Luschnig, C. (2003). PROPORZ1, a putative Arabidopsis transcriptional adaptor protein, mediates auxin and cytokinin signals in the control of cell proliferation. Curr. Biol. **13:** 837–842.

Slewinski, T.L. (2013). Using evolution as a guide to engineer kranz-type c4 photosynthesis. Front. Plant Sci. **4:** 212.

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

Stein, H., Honig, A., Miller, G., Erster, O., Eilenberg, H., Csonka, L.N., Szabados, L., Koncz, C., and Zilberstein, A. (2011). Elevation of free proline and proline-rich protein levels by simultaneous manipulations of proline biosynthesis and degradation in plants. Plant Sci. **181:** 140–50.

Streatfield, S.J., Weber, A., Kinsman, E.A., Häusler, R.E., Li, J., Post-Beittenmiller, D., Kaiser, W.M., Pyke, K.A., Flügge, U.I., and Chory, J. (1999). The phosphoenolpyruvate/phosphate translocator is required for phenolic metabolism, palisade cell development, and plastid-dependent nuclear gene expression. Plant Cell **11:** 1609–1622.

Sud, R.M., and Dengler, N.G. (2000). Cell lineage of vein formation in variegated leaves of the C-4 grass *Stenotaphrum secundatum*. Ann. Bot. (Lond.) **86:** 99–112.

Sugimoto-Shirasu, K., and Roberts, K. (2003). "Big it up": endoreduplication and cell-size control in plants. Curr. Opin. Plant Biol. **6:** 544–553.

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics **22:** 1540–1542.

Tobin, E.M., and Silverthorne, J. (1985). Light regulation of gene-expression in higher plants. Annu. Rev. Plant Biol. **36:** 569–593.

Tolley, B.J., Woodfield, H., Wanchana, S., Bruskiewich, R., and Hibberd, J.M. (2012). Light-regulated and cell-specific methylation of the maize PEPC promoter. J. Exp. Bot. **63:** 1381–1390.

Traas, J., Hülskamp, M., Gendreau, E., and Höfte, H. (1998). Endoreduplication and development: rule without dividing? Curr. Opin. Plant Biol. **1:** 498–503.

Tronconi, M.A., Gerrard Wheeler, M.C., Maurino, V.G., Drincovich, M.F., and Andreo, C.S. (2010). NAD-malic enzymes of *Arabidopsis thaliana* display distinct kinetic mechanisms that support differences in physiological control. Biochem. J. **430:** 295–303.

Vandepoele, K., Raes, J., De Veylder, L., Rouzé, P., Rombauts, S., and Inzé, D. (2002). Genome-wide analysis of core cell cycle genes in Arabidopsis. Plant Cell **14:** 903–916.

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. Plant J. **75:** 656–670.

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

Wheeler, M.C.G., Tronconi, M.A., Drincovich, M.F., Andreo, C.S., Flügge, U.I., and Maurino, V.G. (2005). A comprehensive analysis of the NADP-malic enzyme gene family of Arabidopsis. Plant Physiol. **139:** 39–51.

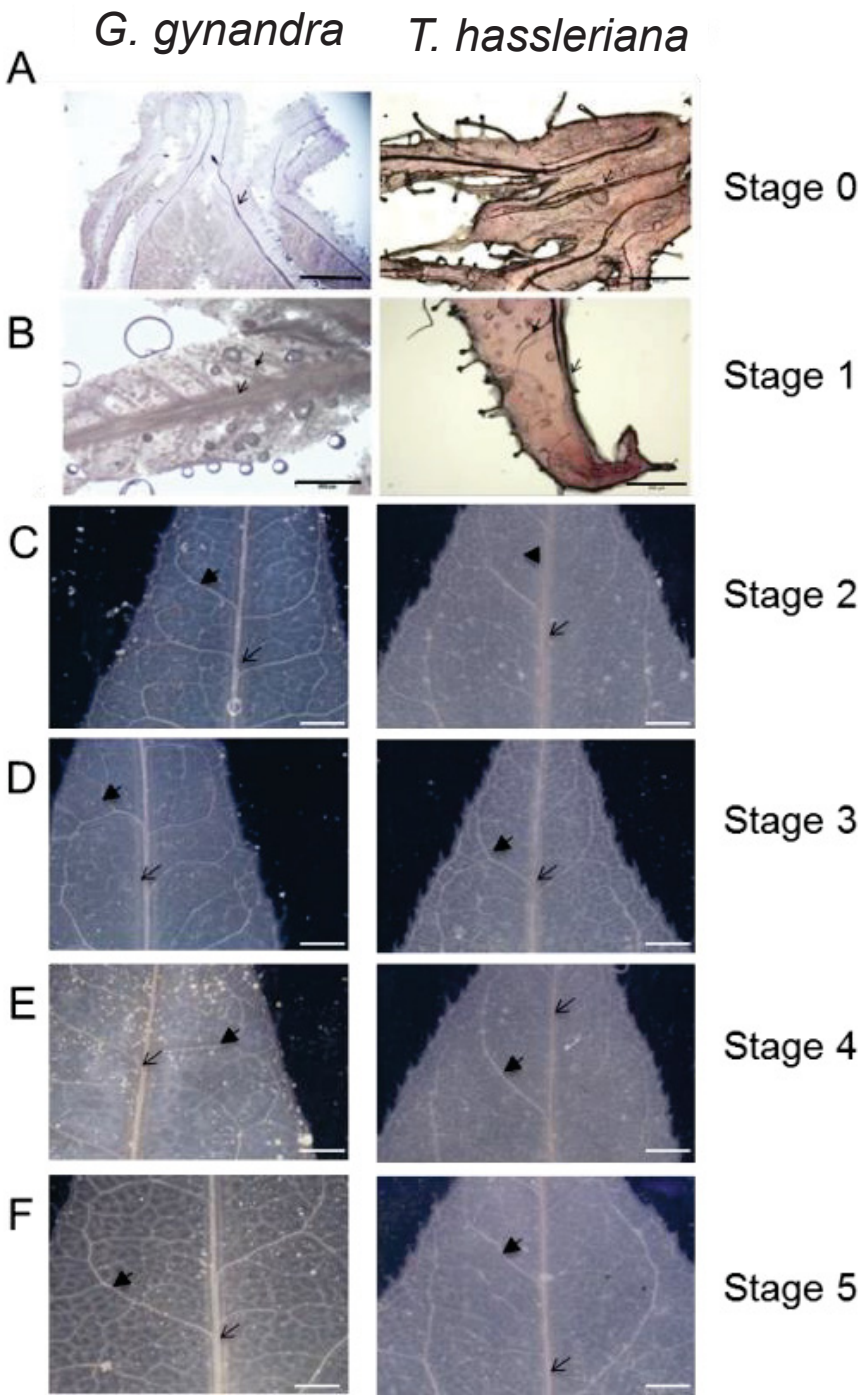
Xu, W., Purugganan, M.M., Polisensky, D.H., Antosiewicz, D.M., Fry, S.C., and Braam, J. (1995). Arabidopsis TCH4, regulated by hormones and the environment, encodes a xyloglucan endotransglycosylase. Plant Cell **7:** 1555–1567.

Yamaguchi, M., Ohtani, M., Mitsuda, N., Kubo, M., Ohme-Takagi, M., Fukuda, H., and Demura, T. (2010). VND-INTERACTING2, a NAC domain transcription factor, negatively regulates xylem vessel formation in Arabidopsis. Plant Cell **22:** 1249–1263.

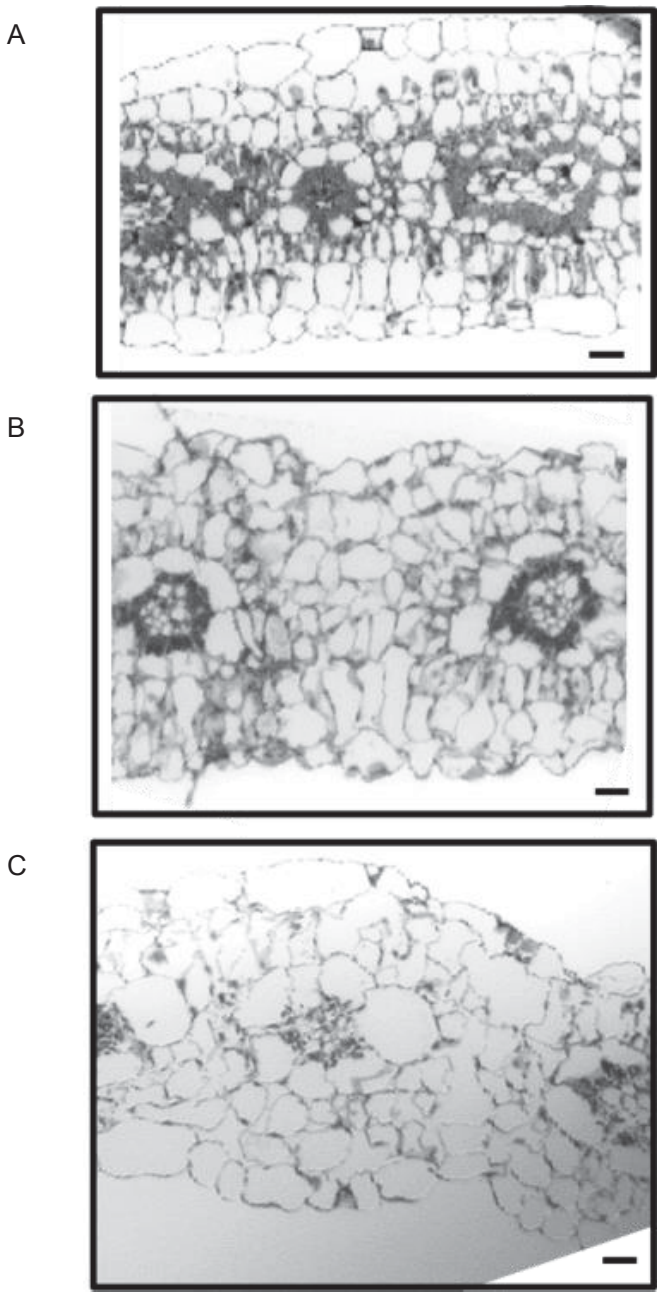
Yekutieli, D., and Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. J. Stat. Plan. Inference **82:** 171–196.

Zhiponova, M.K., et al. (2013). Brassinosteroid production and signaling differentially control cell division and expansion in the leaf. New Phytol. **197:** 490–502.

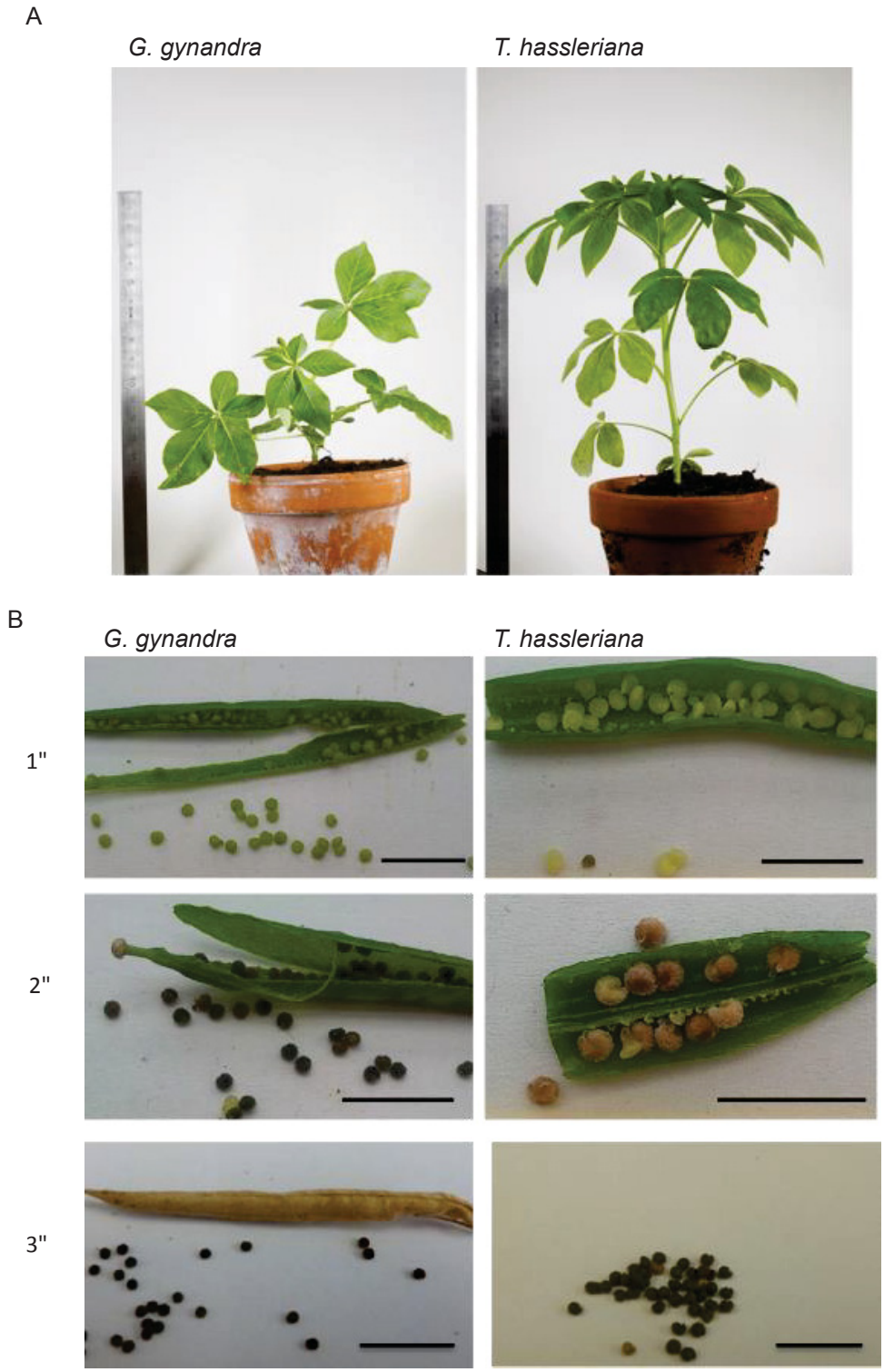
Zhu, X.G., Long, S.P., and Ort, D.R. (2008). What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? Curr. Opin. Biotechnol. **19:** 153–159.



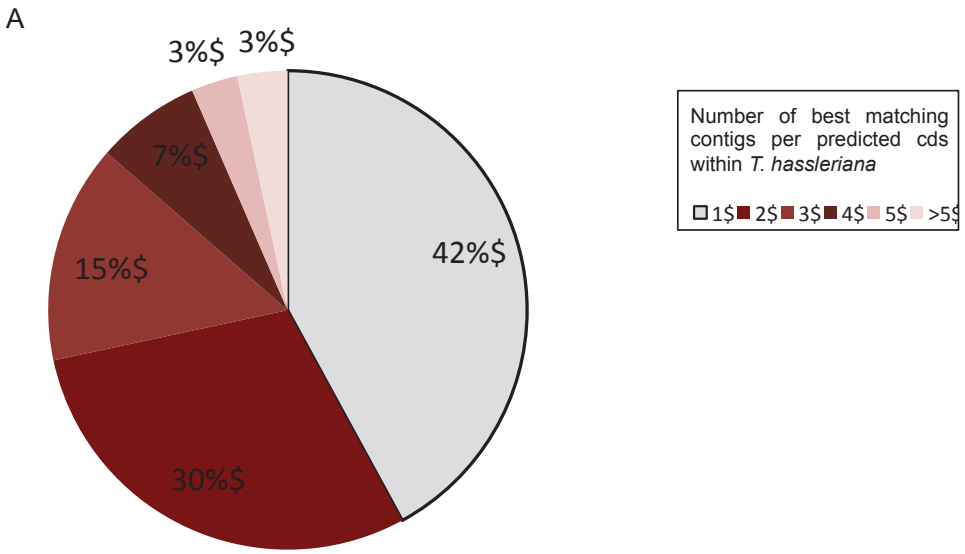
Supplemental Figure 1. Venation patterning during leaf development of *G. gynandra* and *T. hassleriana*.
(A-B) Cleared safranine stained leaves of stage 0 and 1 (n=3; scale bar 0.5 mm)
(C-F) Cleared leaves of stage 2, 3, 4 and 5 respectively (n=3; scale bar 1 mm)
Open arrows indicate the midvein (1°) and closed arrows the secondary vein (2°) localization



Supplemental Figure 2. *G. gynandra* cotyledon anatomy two, four and six days after germination (DAG). Semi-thin cross sections (3 µm) of *G. gynandra* cotyledons after two **(A)**; four **(B)**; six **(C)** DAG. Cross sections were stained with Toluidine Blue. (Scale bar 10 µm, n=3)

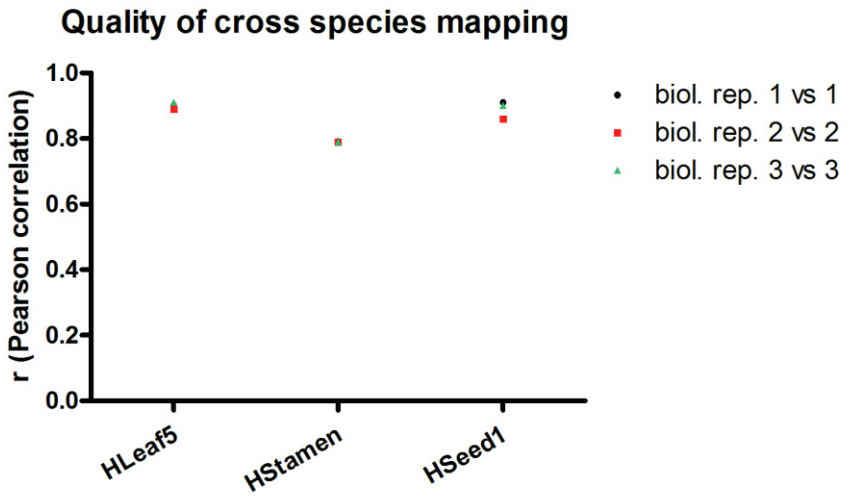


Supplemental Figure 3. Images of tissues harvested for RNA-seq in *G. gynandra* and *T. hassleriana*. (A) Photographic image of *G. gynandra* and *T. hassleriana* 8-week old plants, from which leaf gradient, stem and root system were harvested **(B) Seed coat development from harvested developmental seed gradient. (1)** young seed **(2)** semi-mature seed **(3)** mature seed. (Scale bar = 1cm)

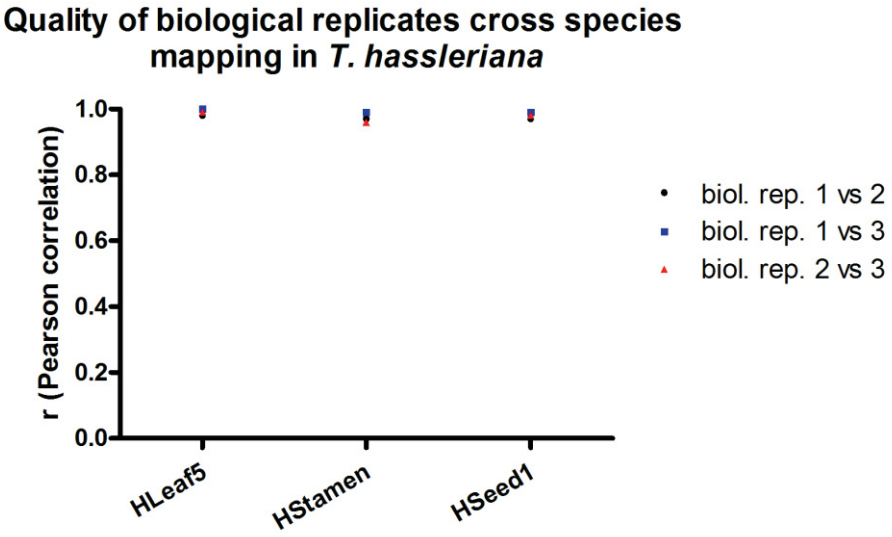


Supplemental Figure 4. Quality assessment of Velvet/OASES assembled *T. hassleriana* contigs against predicted corresponding cds from *T. hassleriana* genome.
(A) Percentage of contig number per predicted cds (Cheng et al., 2013) showing redundancy in assembled contigs.
(B) ClustalW alignment of fragmented contig (top) with corresponding cds (below).
(C) ClustalW alignment of fused contig (top) with corresponding cds (below).

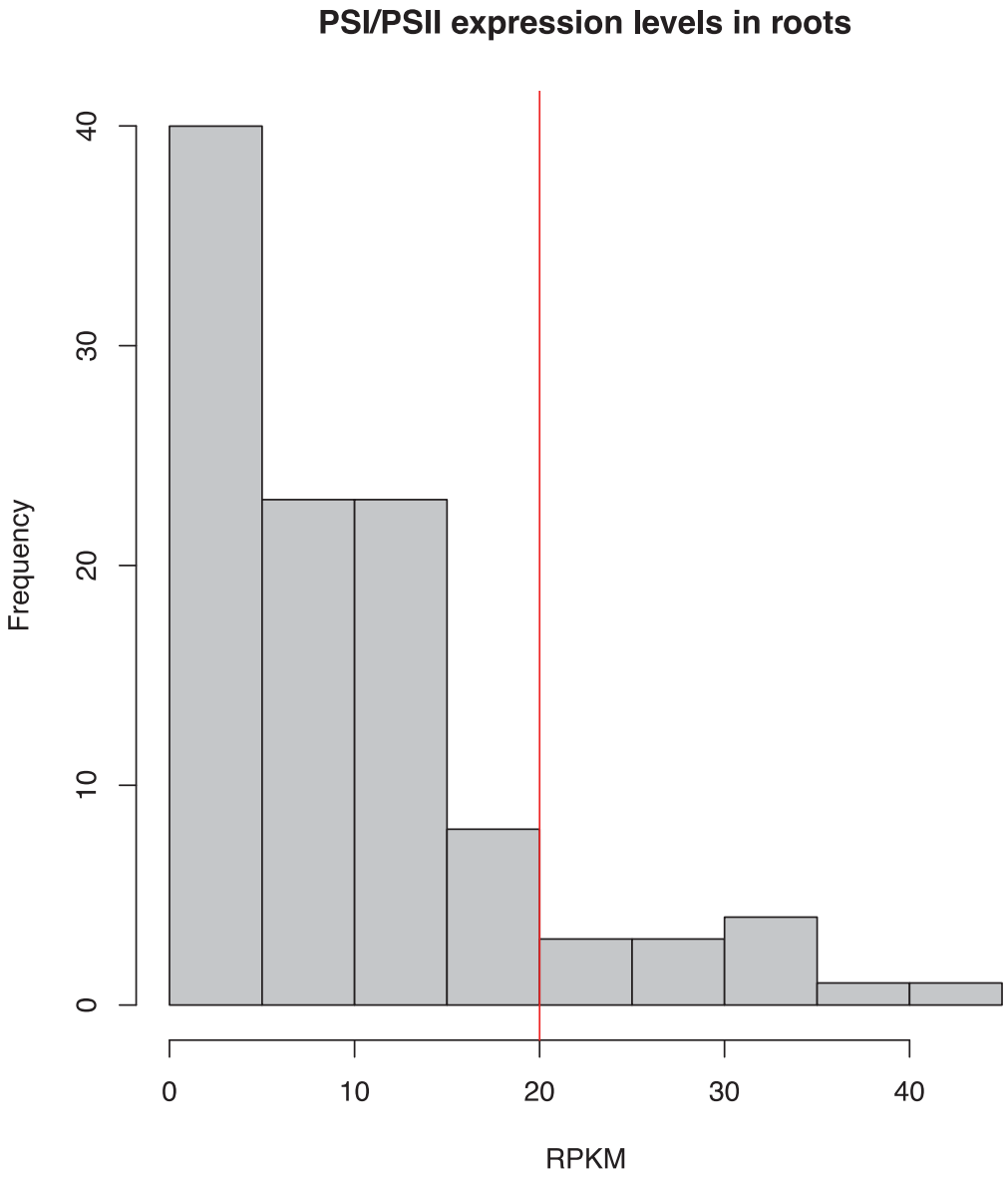
A



B

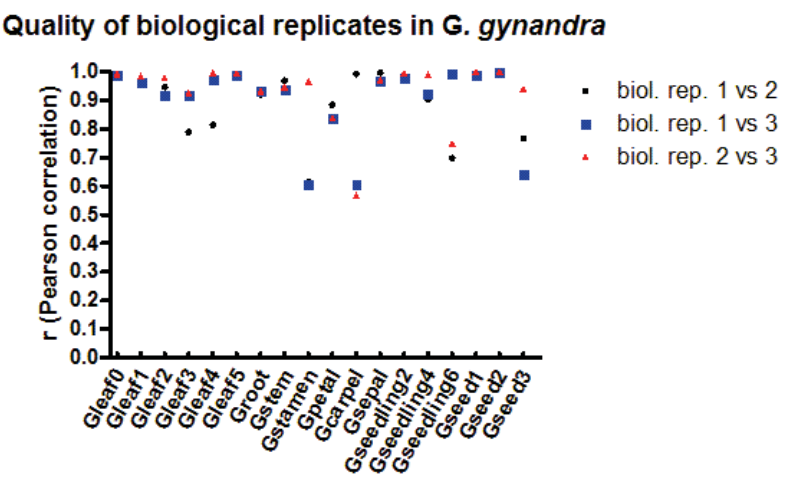


Supplemental Figure 5. Quality assessment of the biological replicates of *T. hassleriana* libraries mapped to *A. thaliana* and mapping similarity of *T. hassleriana* libraries mapped to *A. thaliana* and to its own cds.
(A) Pair-wise Pearson's correlation (*r*) was calculated for all three pairs of biological replicates for each tissue in *T. hassleriana* mapped to *A. thaliana*. **(B)** Pair-wise Pearson's correlation (*r*) between leaf 5, stamen and seed 1 in (n=3) of *T. hassleriana* mapped to its own coding sequence and *A. thaliana*.

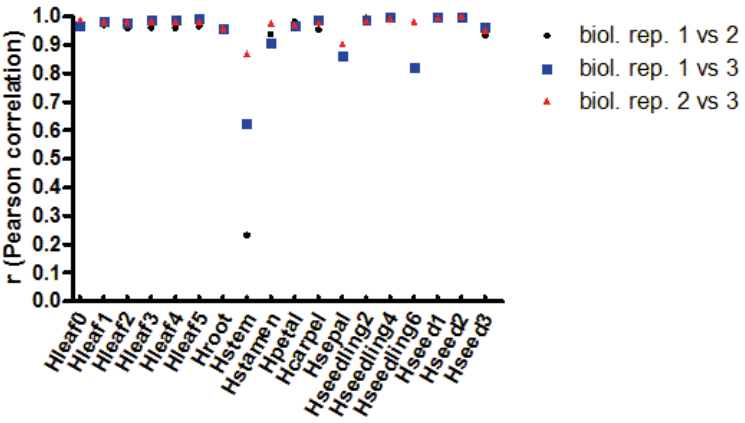


Supplemental Figure 6. Determination of base line gene expression via a histogram of photosystem (PS) I and II transcript abundances reads per mappable million (RPKM) in the *G. gynandra* root.
Y- axis shows frequency and Y- axis depicts RPKM level of PSI and PSII transcript abundance. Red line indicates where threshhold of base line expression was set.

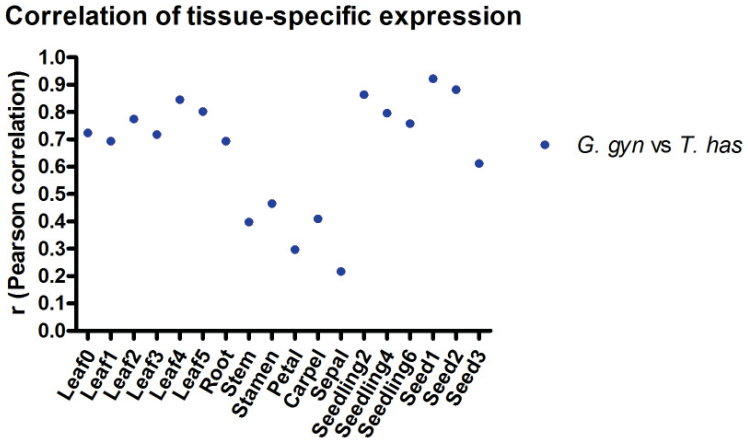
A



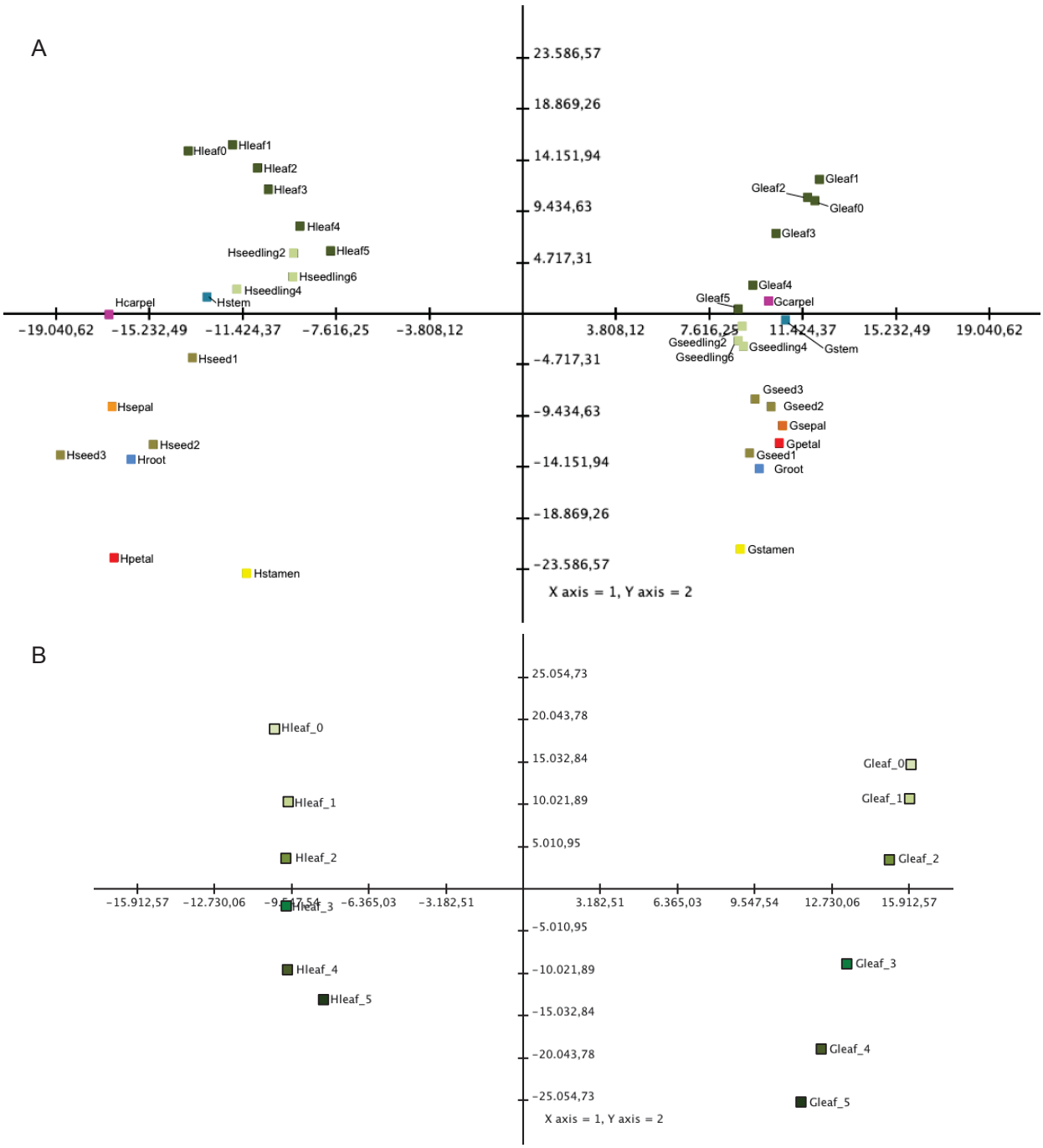
B



C

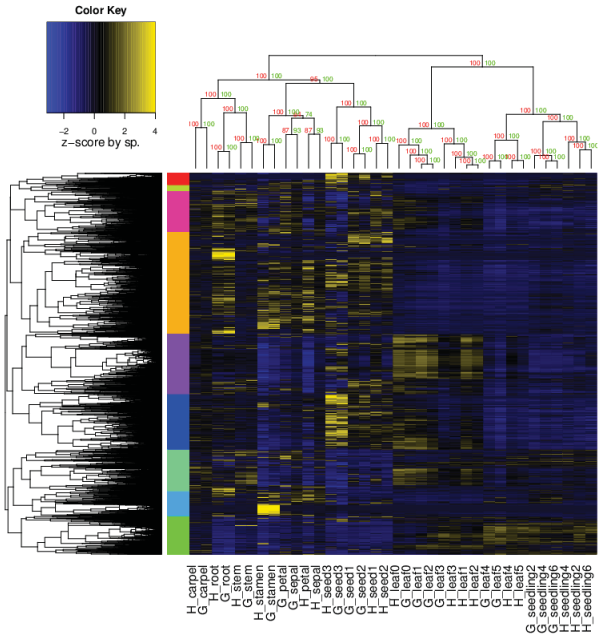


Supplemental Figure 7. Quality assessment of the biological replicates within each species and tissue similarity between *G. gynandra* and *T. hassleriana*. (A) Pair-wise Pearson's correlation (*r*) was calculated for all three pairs of biological replicates for each tissue (n=3) in *G. gynandra*. **(B)** Pair-wise Pearson's correlation (*r*) was calculated for all three pairs of biological replicates for each tissue (n=3) in *T. hassleriana*. **(C)** Pair-wise Pearson‘s correlation between individual tissues of *T. hassleriana* and *G. gynandra*.

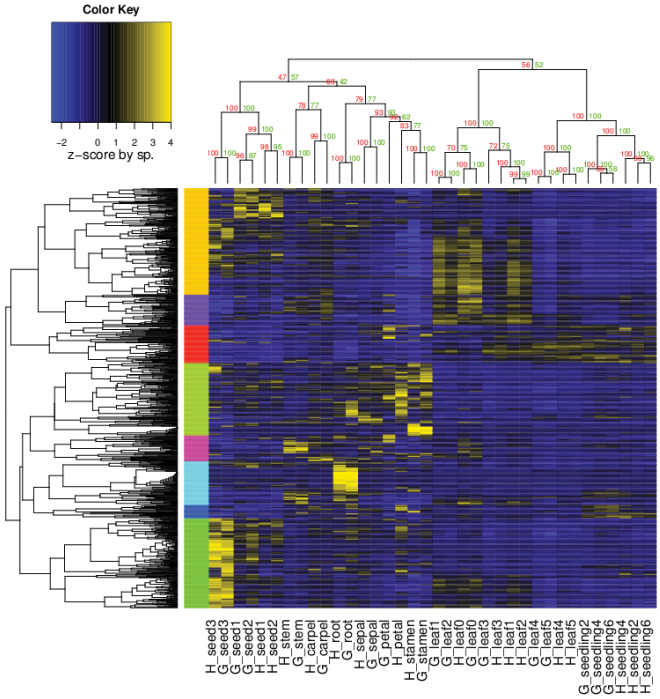


Supplemental Figure 8. Principle component analysis between *G. gynandra* and *T. hassleriana*.
(A) Plot shows all averaged tissues from *G. gynandra* (G) and *T. hassleriana* (H) sequenced (n=3). The first component describes 15% of all data variablility seperating both species. The second component (14%) separates samples by tissue identity within each species. Tissues are indicated by color key (left).
(B) Averaged leaf gradient samples (n=3) from *G. gynandra* (G) and *T. hassleriana* (H) were analysed. First component decribes 44 % and second component describes 29% of variability.

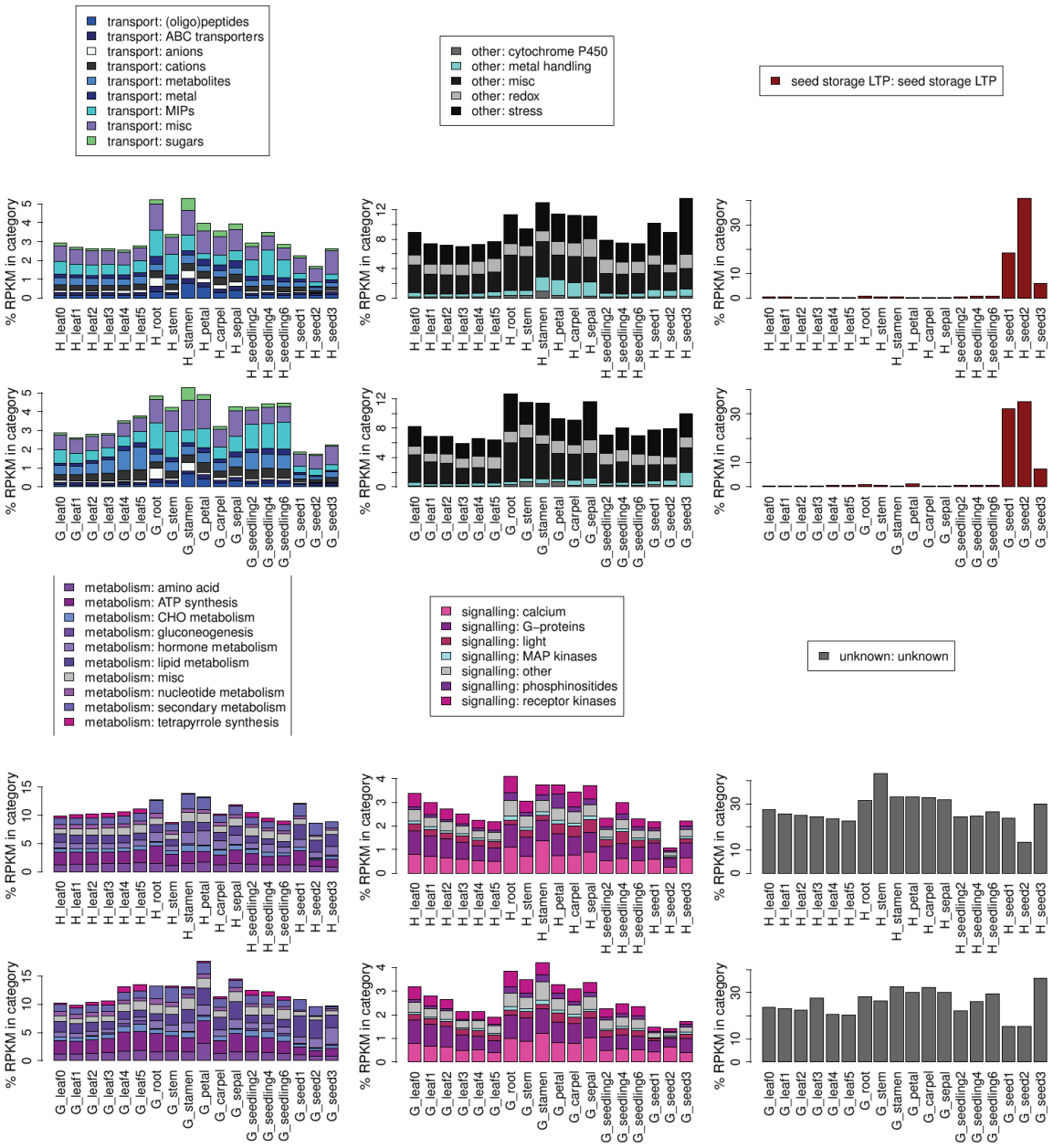
A



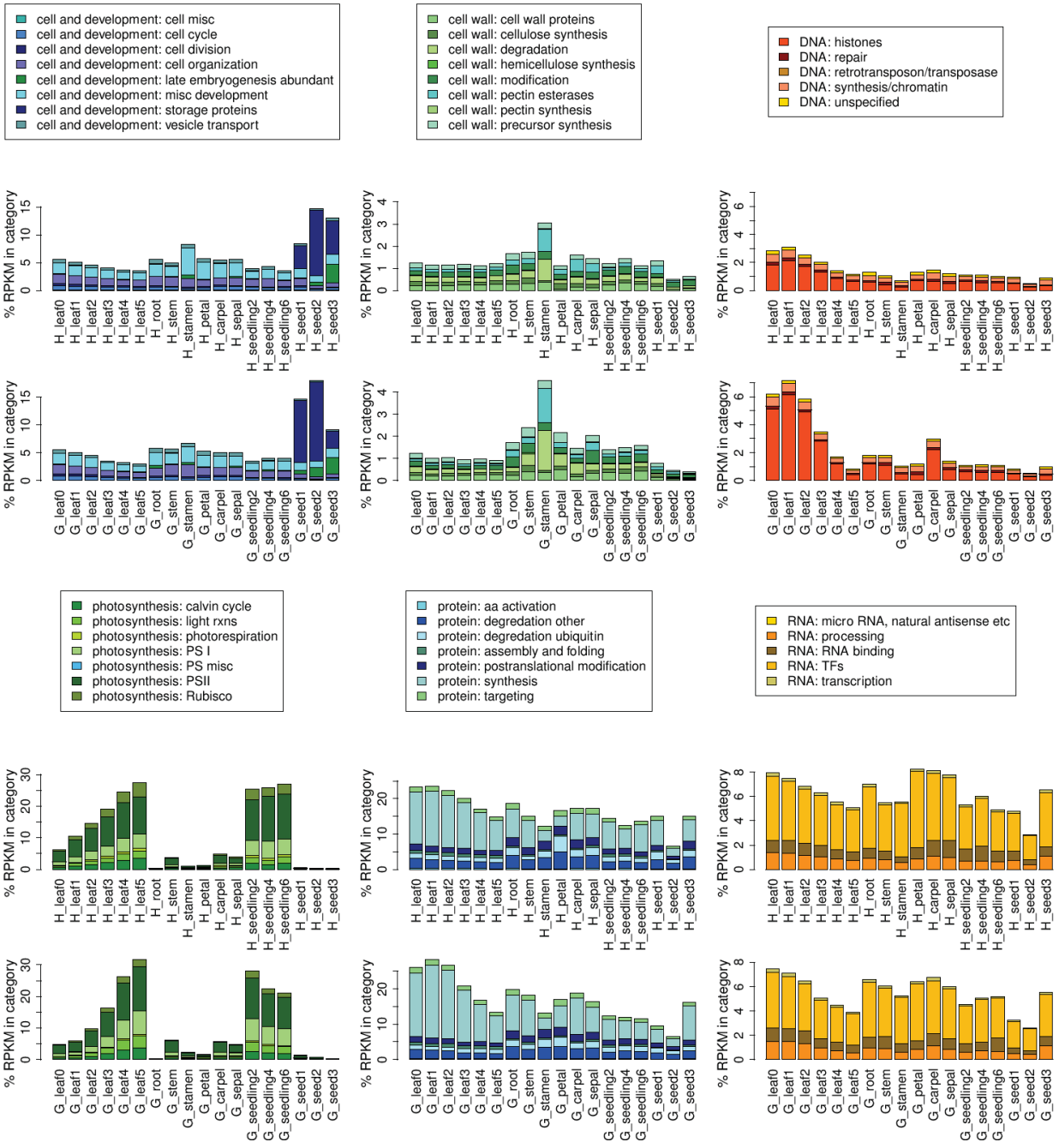
B



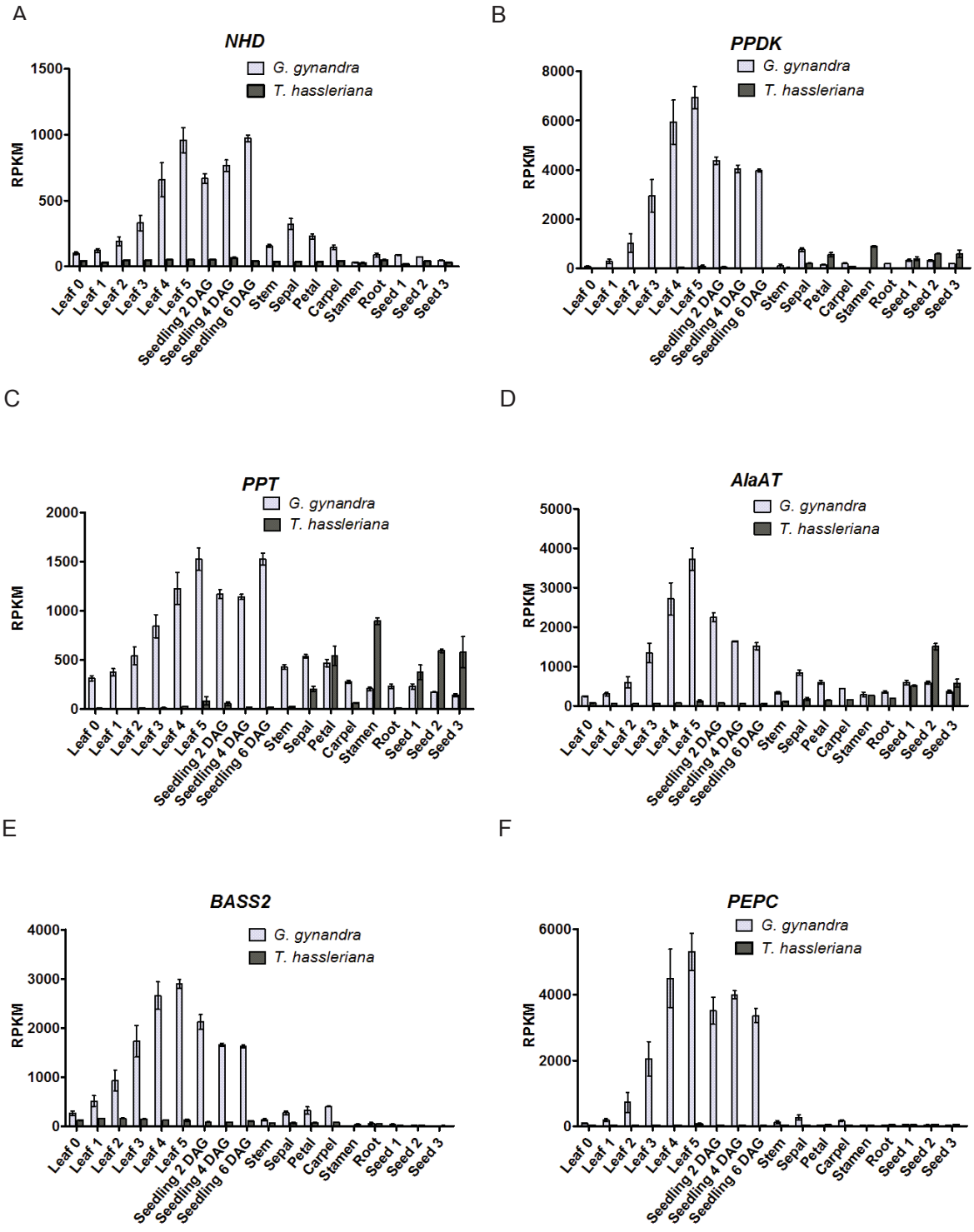
Supplemental Figure 9. Hierarchical cluster analysis with bootstrapped samples of *G. gynandra* and *T. hassleriana*. Numbers above the nodes show the approximately unbiased p-value (red) and the bootstrap probability (green). Blue is lowest expression and yellow highest expression.'Left-hand vertical bars denote major clusters in the dendrogram by color. **(A)** Clustering of all over 20 RPKM expressed genes in all averaged samples (n=3). Sample averages were clustered as species scaled Z-scores with Pearson‘s Correlation. **(B)** Hierarchical Clustering of all transcriptional regulators expressed in all tissues sequenced in *G. gynandra* and *T. hassleriana*. Sample averages (n=3) were clustered as species-scaled Z-scores with Pearson‘s Correlation.



Supplemental Figure 11.1. Transcriptional investment at secondary Mapman category of each tissue compared in both species (Part 1). Distribution of the Mapman categories in each tissue in *G. gynandra* and *T. hassleriana*. Plot shows percent of average RPKMs of the 12 customized secondary Mapman bins for each tissue.

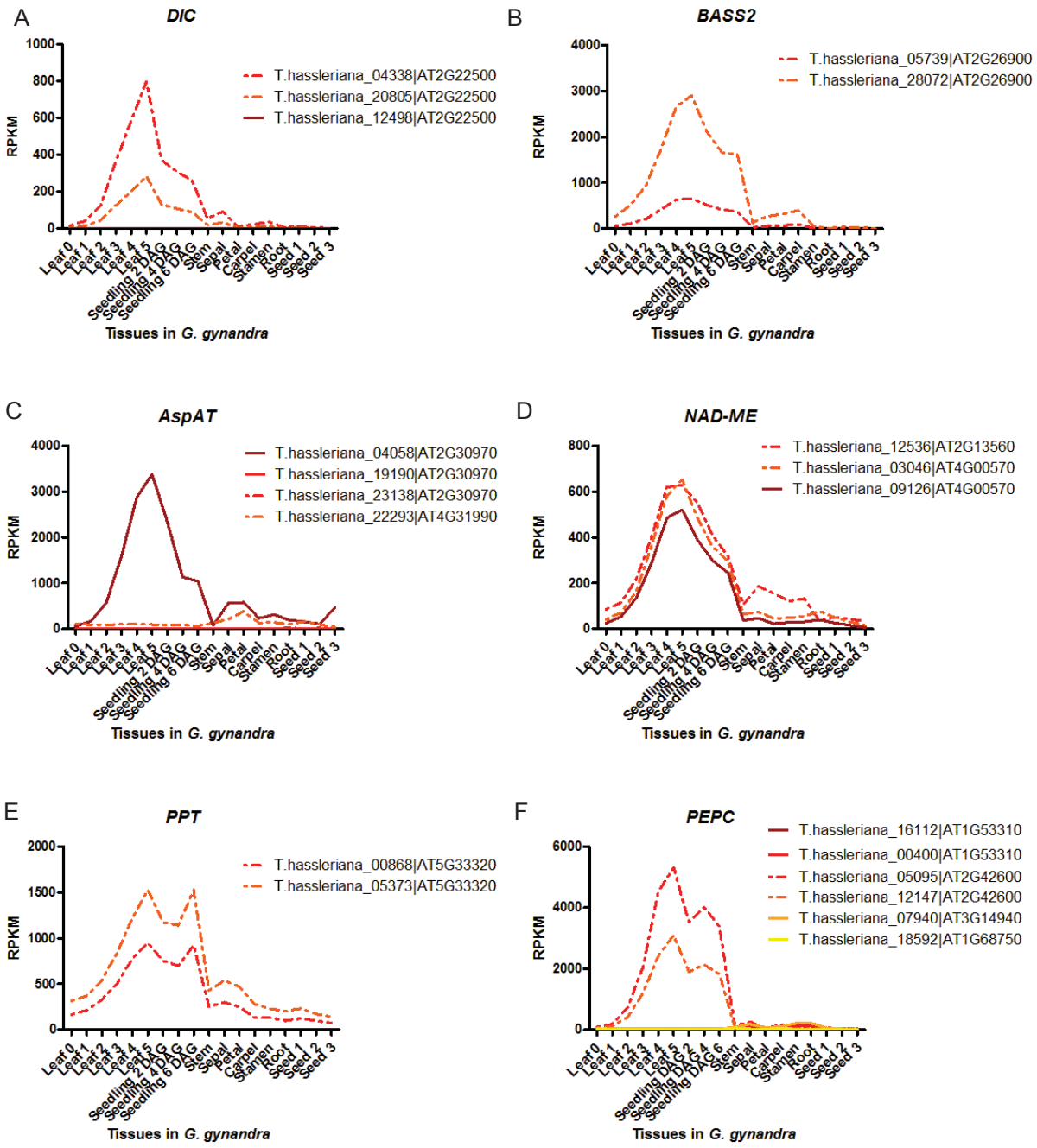


Supplemental Figure 11.2. Transcriptional investment at secondary Mapman category of each tissue compared in both species (Part 2). Distribution of the Mapman categories in each tissue in *G. gynandra* and *T. hassleriana*. Plot shows percent of average RPKMs of the 12 customized secondary Mapman bins for each tissue.

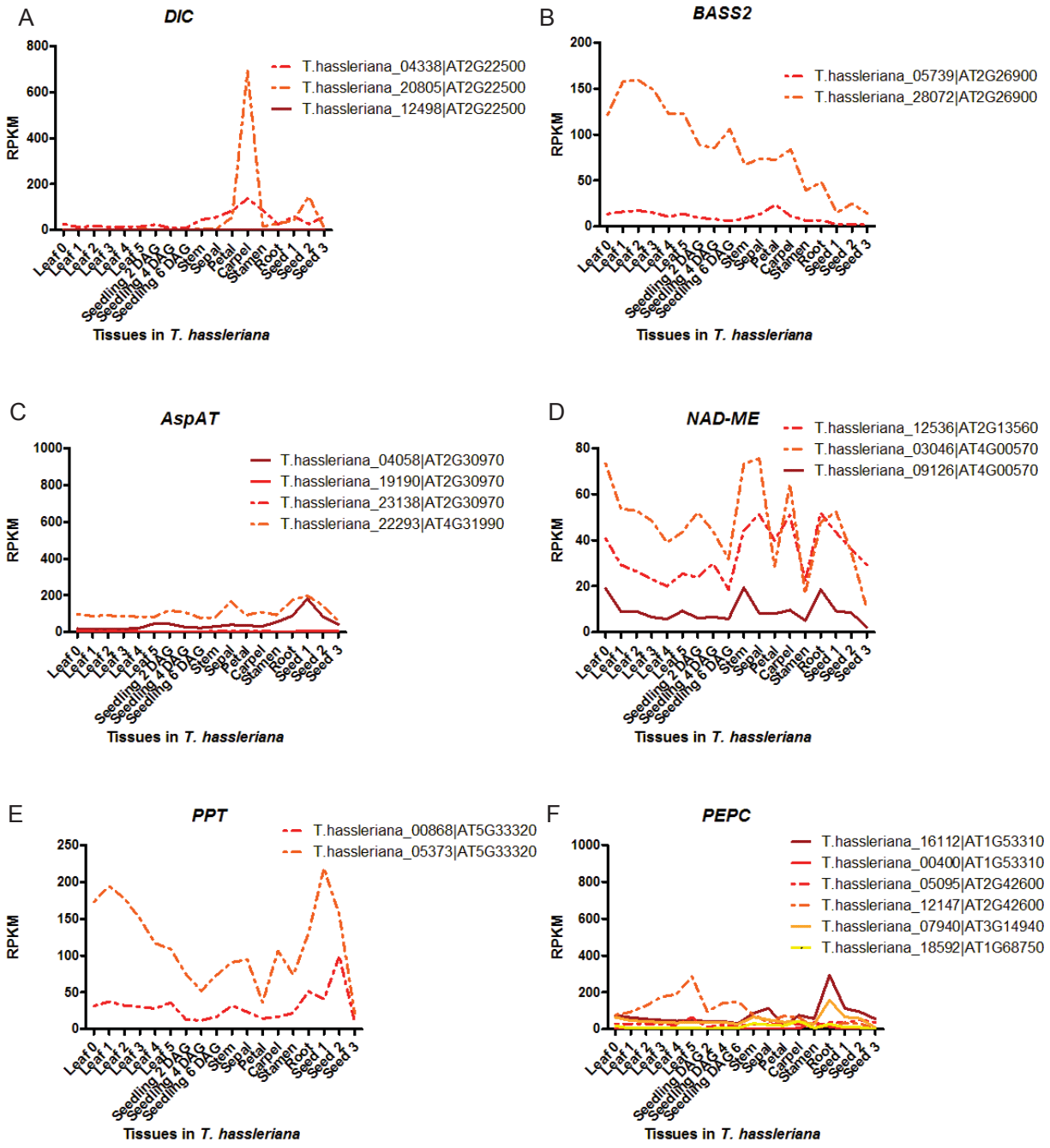


Supplemental Figure 12. Comparison of gene expression dynamics within the leaf gradient of both species.

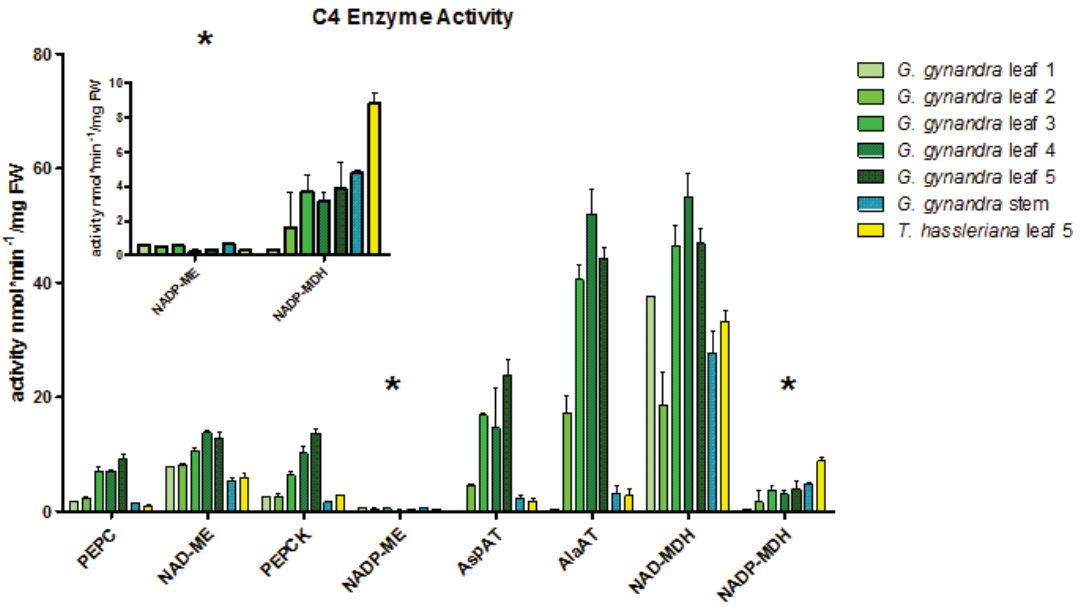
(A-F) Average expression pattern of highest abundant putative ortholog of C_4 cycle genes (*NHD, PPDK, PPT, AlaAT, BASS2, PEPC*) in photo- and heterotrophic tissues in *G. gynandra* (light grey) and *T. hassleriana* (dark grey); (n=3 ± SE, standard error)



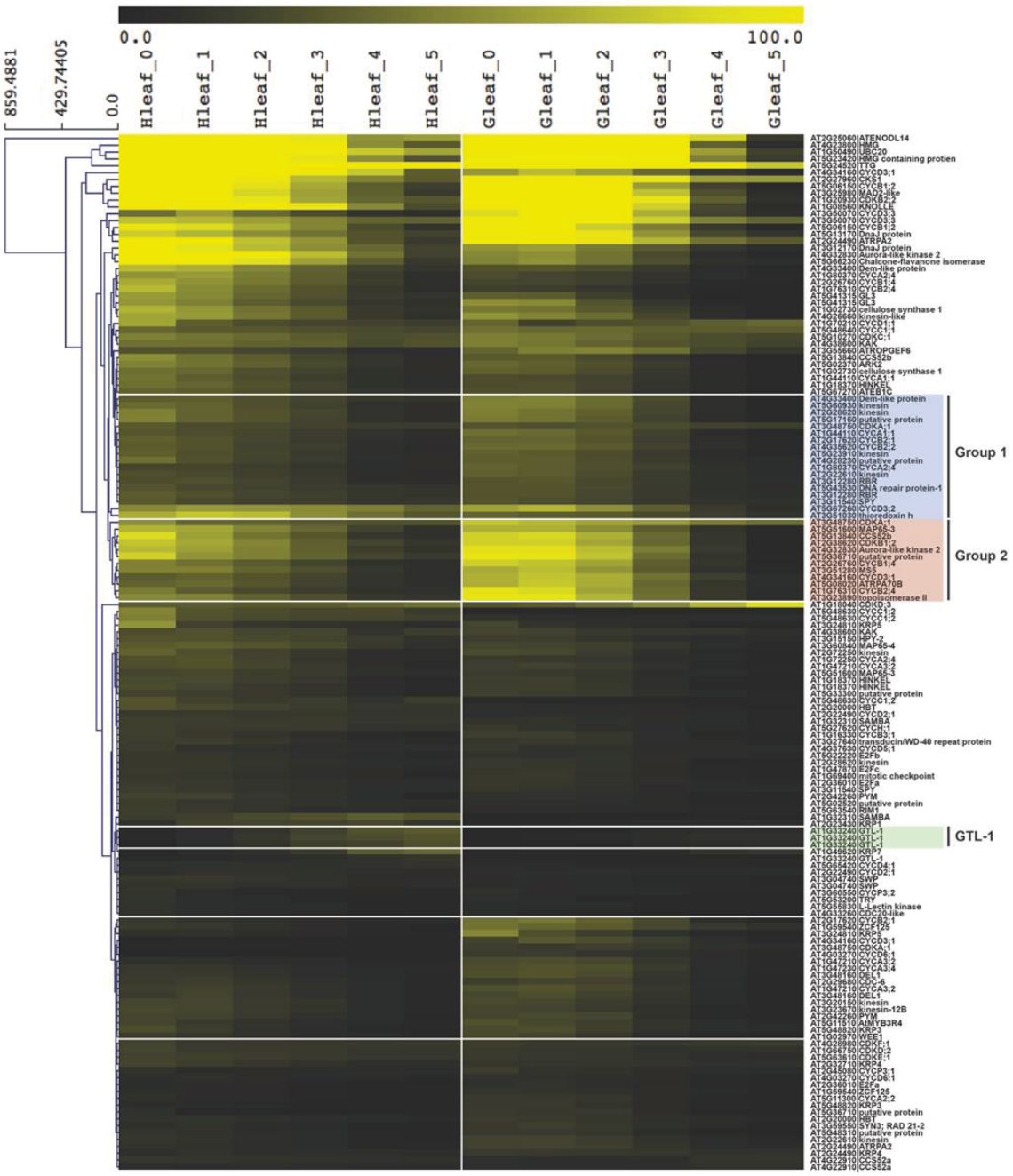
Supplemental Figure online 13. Plot of all C_4 gene putative orthologs expression pattern (RPKM) in *G. gynandra,* that were annotated as C_4 genes with AGI identifier and respective *T. hassleriana* ID. **(A-F)** Average expression pattern of putative ortholog of C_4 cycle genes (*DIC, BASS2, AspAT, NAD-ME, PPT, PEPC*) in photo- and heterotrophic tissues in *G. gynandra* (n=3).



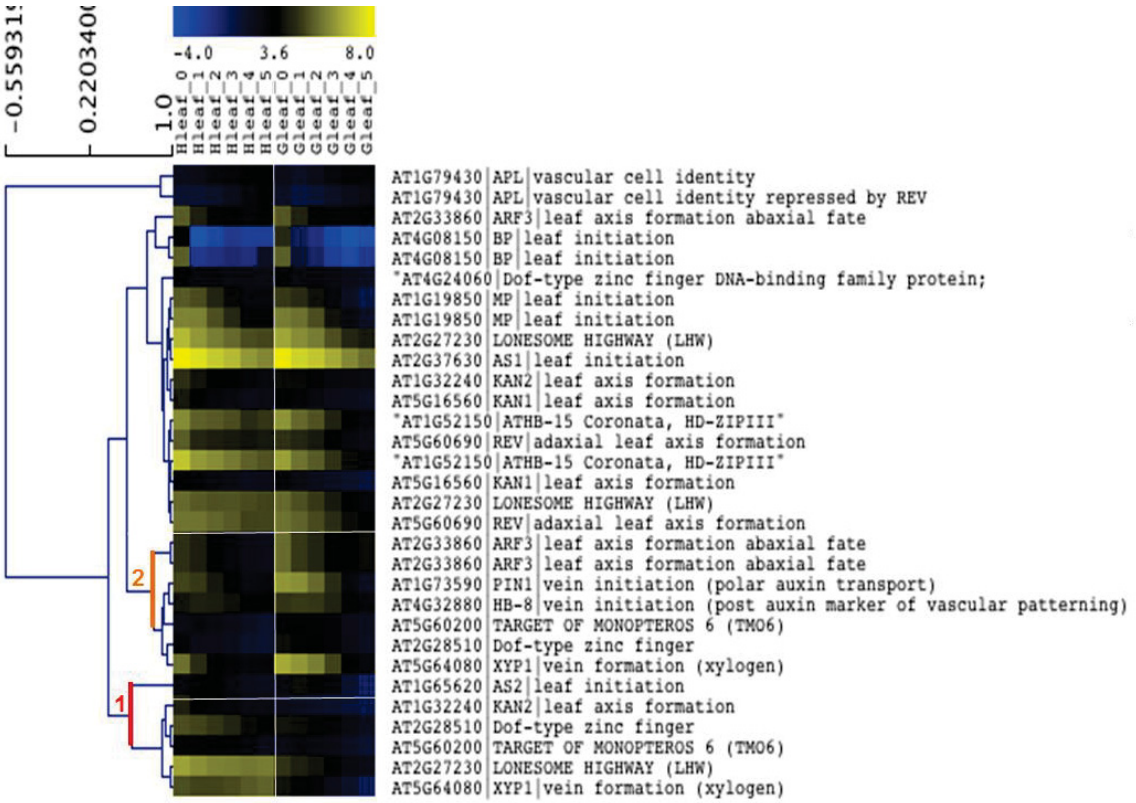
Supplemental Figure online 14. Plot of all C_4 gene putative orthologs expression pattern (RPKM) in *T. hassleriana,* that were annotated as C_4 genes with AGI identifier and respective *T. hassleriana* ID. **(A-F)** Average expression pattern of putative ortholog of C_4 cycle genes (*DIC, BASS2, AspAT, NAD-ME, PPT, PEPC*) in photo- and heterotrophic tissues in *T. hassleriana* (n=3).



Supplemental Figure 15. Enzyme activity measurement of soluble C_4 cycle enzymes. Enzyme activities of PEPC, NAD-ME, PEPCK, NADP-ME, AspAT, AlaAT, NAD-MDH and NADP-MDH were measured along the developing *G. gynandra* leaf (stage 1-5) with the mature *T. hassleriana* leaf (stage 5) as C_3 control. (FW: fresh weight; n=3 ±SE, standard error; biological replicates with each 3 technical replicates)

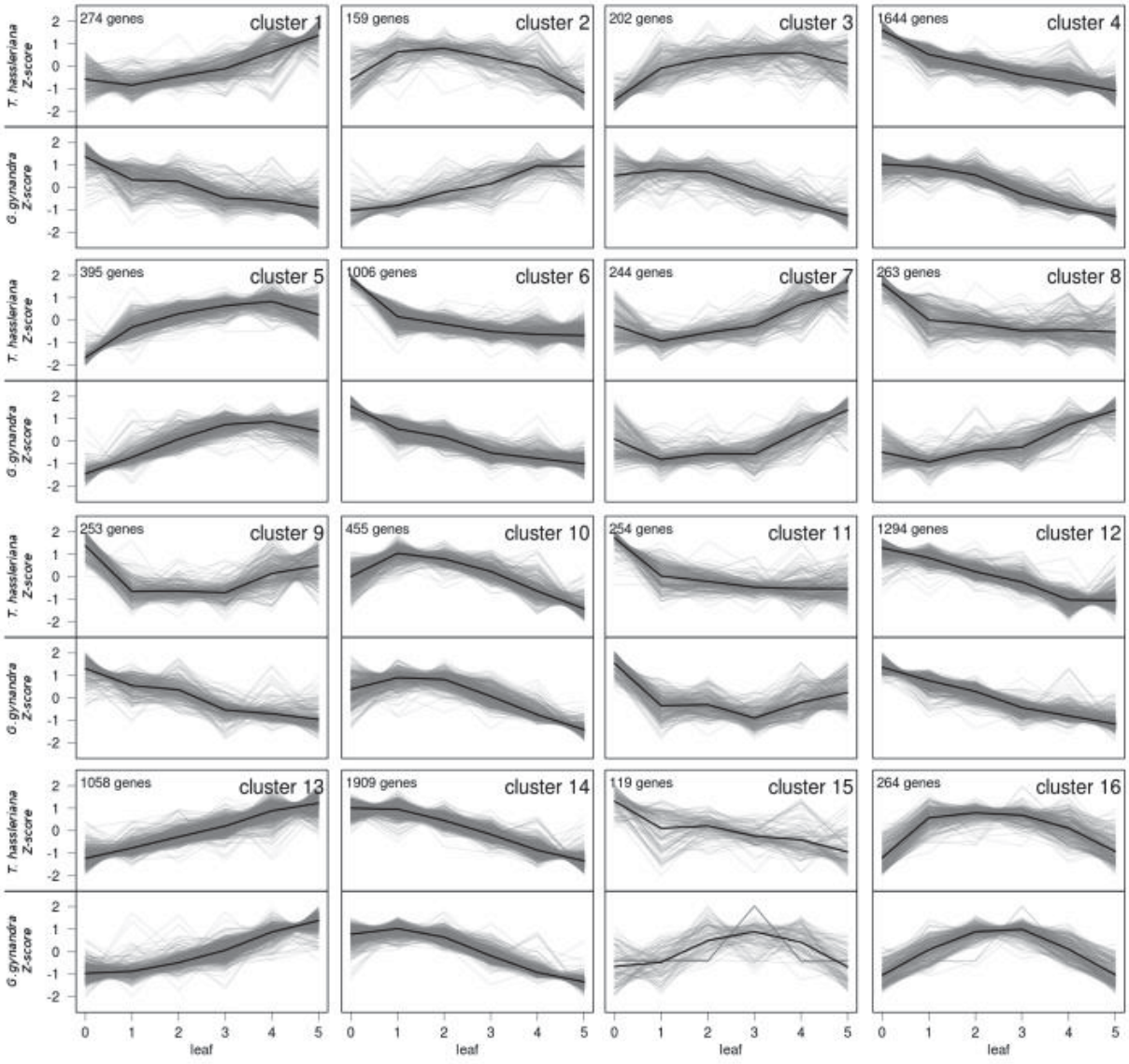


Supplemental Figure 16. Hierarchical clustering of average RPKM with Euclidean distance of core cell cycle genes in *T. hassleriana* and *G. gynandra*. Core cell cycle genes were extracted from (Vandepoele et al., 2002; Beemster et al., 2005). Deregulated cluster of interest are marked with blue and red boxes. *GTL1* cluster is highlighted with green box.

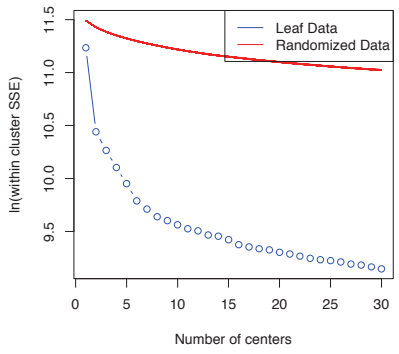


Supplemental Figure 17. Hierarchical clustering with Pearson's correlation of leaf developmental factors. Averaged transcript abundances (RPKM) of leaf gradient sample of transcriptional regulators involved in axial and vasculature fate determination were clustered. Group 1 (orange) and group 2 (red) show genes that are altered between *T. hassleriana* (H) and *G. gynandra* (G).

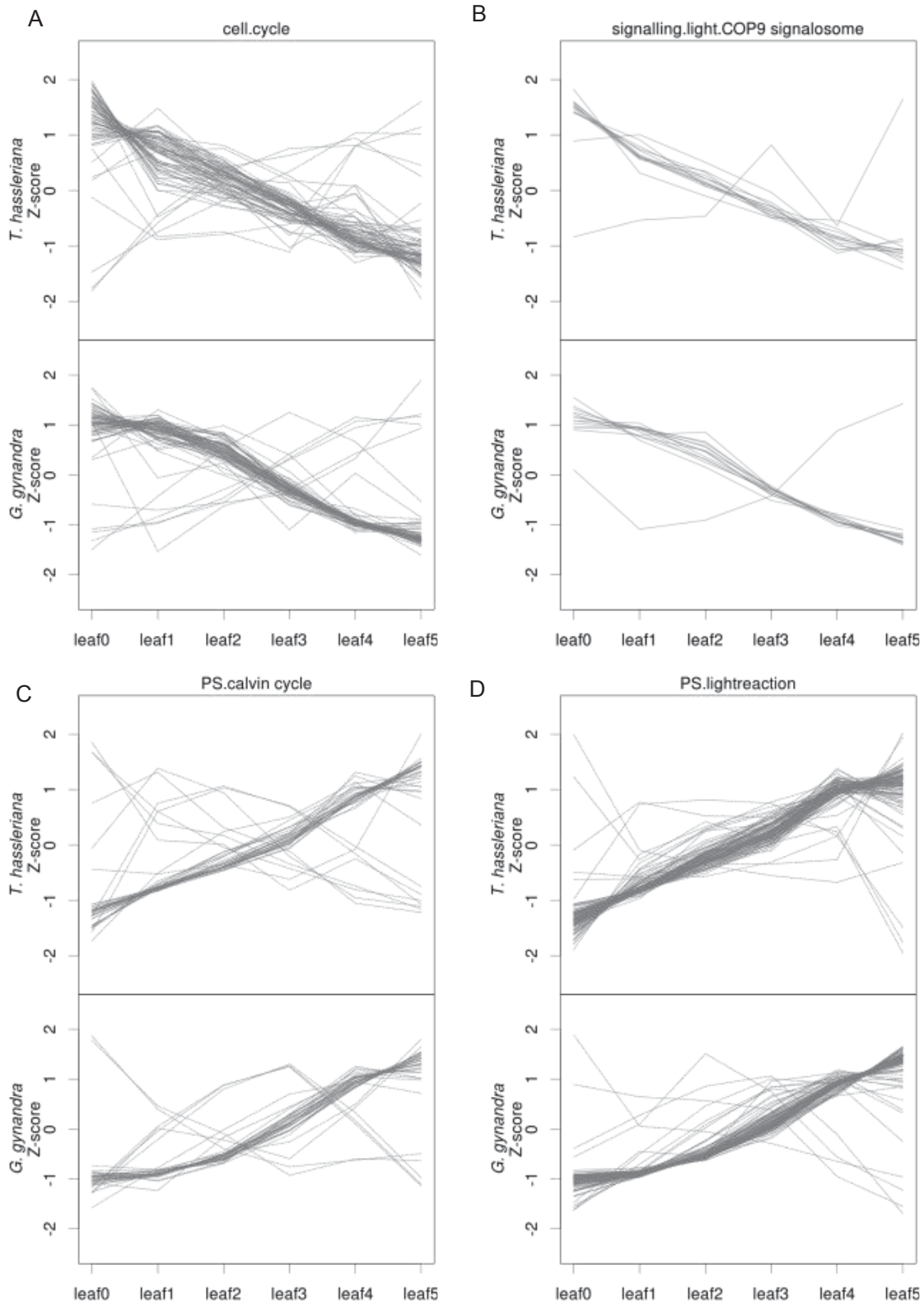
A



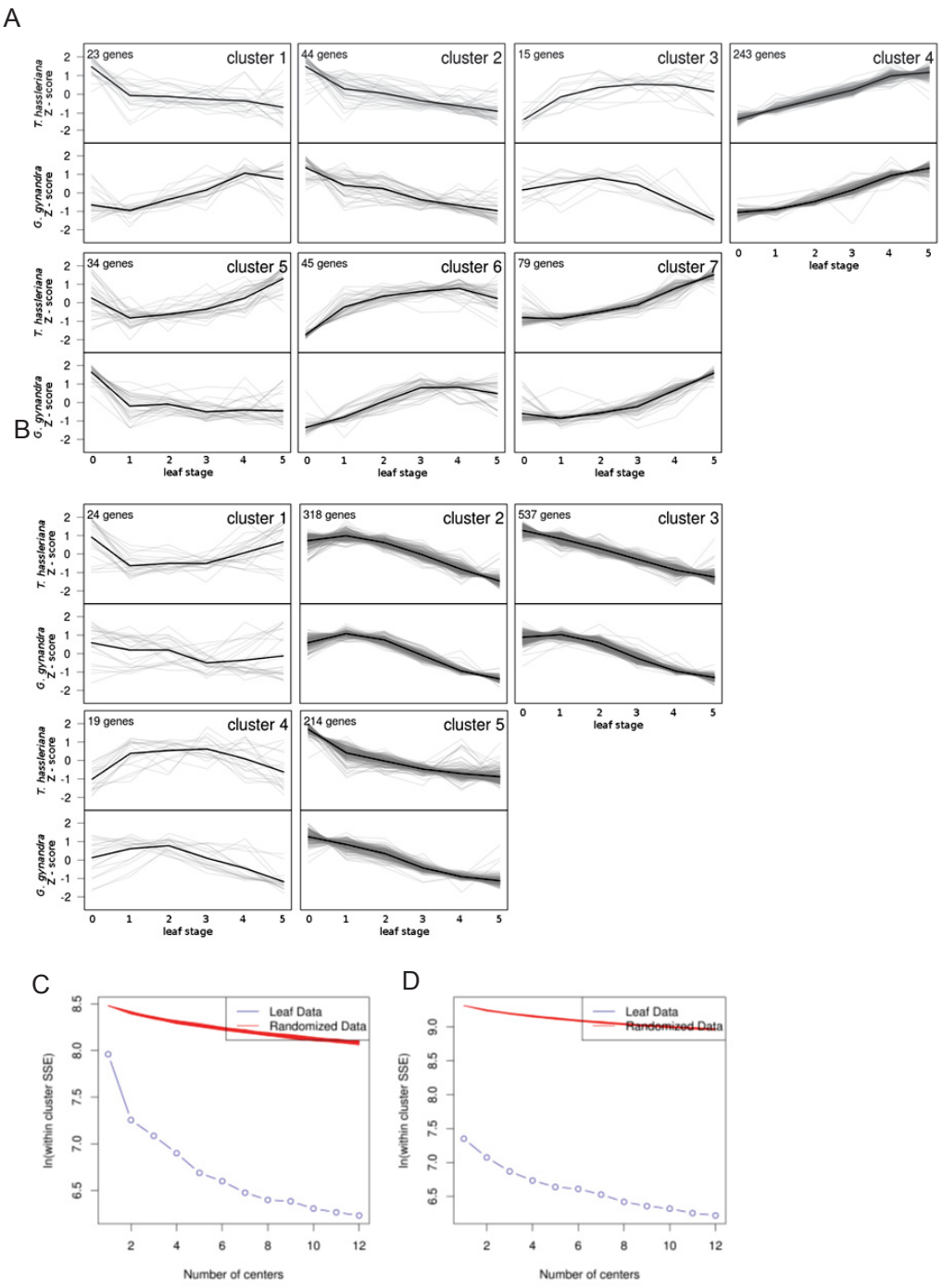
B



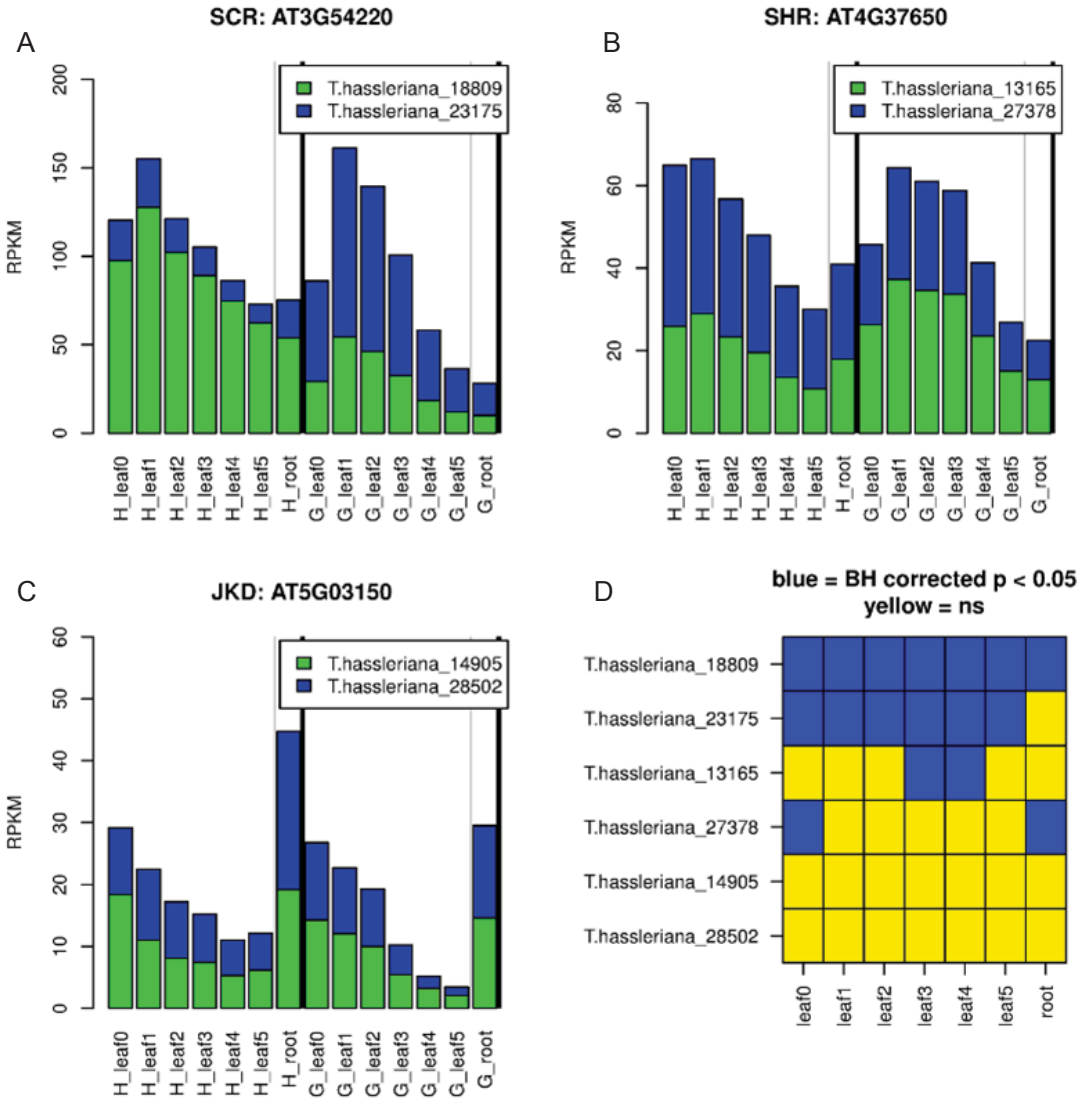
Supplemental Figure 18. *K*-means clustering of leaf gradient expression data and quality assessment. **(A)** *K*-means clustering of transcript abundances (RPKM) of leaf stage averages (*n*=3) between *T. hassleriana* and *G. gynandra* shown as species-scaled *Z*-scores. Size of each cluster is indicated in each cluster box. **(B)** Ln of the sum of the squared euclidean distance (SSE) between each gene and the center of it's cluster across various numbers of clusters calculated with a *K*-means algorithm for the leaf gradient data (blue) compared to the average of 250 scrambled datasets (red).



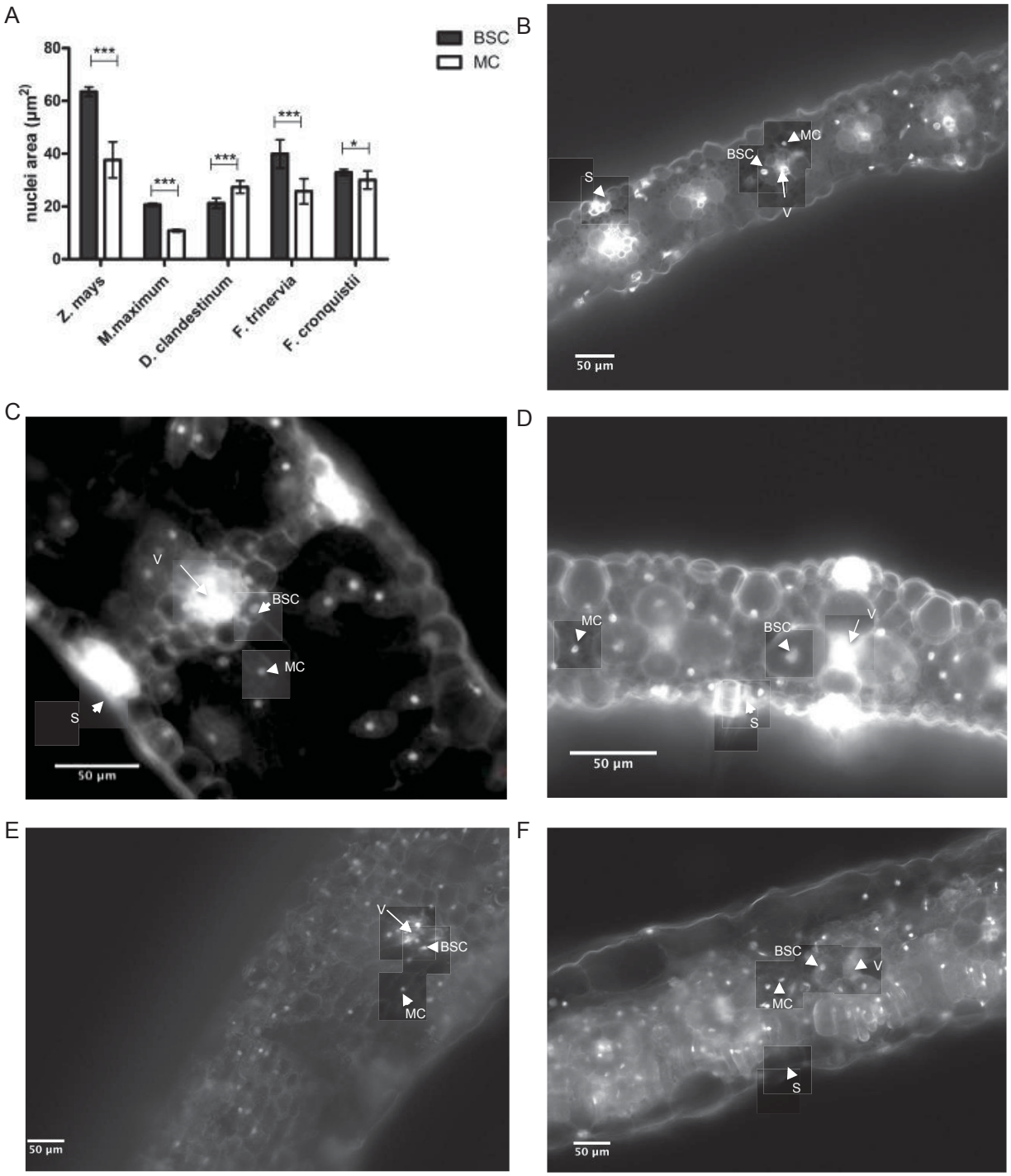
Supplemental Figure 19. Z-score plots of enriched mapman categories in the shifted clusters. Species scaled Z-scores from averaged transcript abundances (RPKM) for each leaf stage per species (n=3). **(A,B)** shifted enriched categories from cluster 4. **(C,D)** shifted enriched categories from cluster 13. Number in brackets are the respective Mapman category bin codes.



Supplemental Figure 20. K-means clustering of genes differentially regulated during the transition from proliferation to enlargement. (A,B) K-means clustering of *T. hassleriana* and *G. gynandra* homologs of gene set that is significantly up-regulated **(A;** p-value<0.05**)** or down-regulated (**B**; p-value<0.05) between day 9 and 10 day in developing *A. thaliana* leaves (Andriankaja et al., 2012). Per species scaled Z-scores from averaged transcript abundances (RPKM) for each leaf stage per species (n=3). **(C,D)** Ln of the sum of the squared Euclidean distance (SSE) between each gene and the center of its clusters across various numbers of clusters calculated with a K-means algorithm for the leaf gradient data (blue) compared to the average of 250 scrambled datasets (red) for **(C)** up- and **(D)** down-regulated.



Supplemental Figure 21. Transcript abundances of *SCARECROW* and *SHORTROOT* homologs in *G. gynandra* (G) and *T. hassleriana* (H) leaf and root.
(A-C) Expression pattern (average RPKM; n=3) of all homologs of *SCARECROW* (***SCR*; A**); *SHORTROOT* (***SHR*; B)** and *JACKDAW* (***JKD*; C)** in both species. **(D)** Dual color map of significant (blue; FWE corrected p-Value<0.05) or non significant (yellow; n.s) expressed transcripts of *SCR*, *SHR* and *JKD*.



Supplemental Figure 22. Nuclei area and images of C_4 and C_3 species.
(A) Quantification of BSC and MC nuclei area of mature leaves of monocotyledonous (*Zea mays*; *Megathyrsus maximus*; *Dichantelium clandestinum*) and dicotyledonous (*Flaveria trinervia*; *Flaveria cronquistii*) C_4 and C_3 species cross sections (error bars ±SD; n=3). Area of nuclei is given as µm² with at least 100 nuclei analyzed per cell type per species. Asterisks indicate statistically significant differences between BSC and MC (*** p-value<0.001; * p-value<0.05). **(B-F)** Microscopic fluorescence images of propidium iodide stained mature leaf cross sections of *Zea mays*,C_4 (**B**); *Dichantelium clandestinum*;C_3 (**C**); *Megathyrsus maximus*, C_4 (**D**); *Flaveria cronquistii*, C_3 (**E**); *Flaveria trinervia,* C_4 (**F**). Scale bar: 50 µm; closed arrows pointing to nuclei of indicated cell type. BSC: bundle sheath cell; MC: mesophyll cell; V: vein; S: stomata.

Supplemental Table 1 online. Velvet/OASES assembly stats from *G. gynandra* and *T. hassleriana* paired end reads. Backmapping of paired end reads was performed with TopHat standard settings. Annotation via blastp against TAIR10 proteome.

	G. gynandra (C_4)	*T. hassleriana* (C_3)
k-mer	31	31
N50 contig	1916	1996
unigenes	59471	52479
total transcripts	176850	163456
Backmapping %	60	63
Annotation of TAIR10 %	86	87

Supplemental Table 2 online. Cross species mapping results. *T. hassleriana* Leaf 5, Seed 1, Stamen (n=3) was mapped to *A. thaliana* via blat in translated protein (A) mode to assess sensitvity of cross species mapping. Results of mapping were normalized as RPKM and collapsed on 1 AGI per multiple identifier in *T. hassleriana* Pearson's correlation *r* values of collapsed *T. hassleriana* Leaf 5, Seed 1 and Stamen (n=3) mapped to *A. thaliana* (B) and to itself were calculated (C).

A

Species	Sample	Total number of cleaned reads	Total number of mapped reads	Mapping efficiency against A.thaliana reference	Number of genes>20 RPKM	Number of genes>1000 RPKM
T. hassleriana	Hleaf5_1	41085063	23502678	57.20492141	5825	151
	Hleaf5_2	26393836	22289304	84.44889936	5675	122
	Hleaf5_3	67907227	43184738	63.59372913	5684	146
	Hstamen_1	46237107	27726175	59.96520284	5923	48
	Hstamen_2	48025041	28220020	58.76105343	5950	47
	Hstamen_3	17855771	14433105	80.83159781	5467	60
	Hseed1_1	38620315	21654259	56.06960741	6253	39
	Hseed1_2	28792149	17462026	60.64856777	6301	48
	Hseed1_3	25372947	14217549	56.03428329	6107	42

B

collapsed expression by mapping to own cds vs to A. thaliana		1vs1	2vs2	3vs3	average
Hleaf5	r	0.90	0.89	0.91	0.90
	r2	0.81	0.80	0.82	0.81
Hstamen	r	0.79	0.79	0.79	0.79
	r2	0.62	0.62	0.62	0.62
Hseed1	r	0.91	0.86	0.9	0.89
	r2	0.83	0.74	0.81	0.79

C

T. hassleriana mapped to *A. thaliana*		1vs2	1vs3	2vs3	average
Hleaf5	r	0.98	1.00	0.98	0.99
	r2	0.97	0.99	0.96	0.97
Hstamen	r	0.97	0.96	0.98	0.97
	r2	0.94	0.92	0.96	0.94
Hseed1	r	0.97	0.99	0.98	0.98
	r2	0.94	0.98	0.96	0.96

Supplemental Table 3 online. Pearson's correlation (r) of each individual replicate per tissue in G. gynandra and T. hassleriana respectively (A). Pearson's correlation between G. gynandra and T. hassleriana individual tissues (B).

A

#	Species	Tissue	Pearson correlation *r* between biological replicates 1 vs 2	1 vs 3	2 vs 3
1	*G. gynandra*	**Gleaf0**	0.98	0.99	0.99
2		**Gleaf1**	0.97	0.96	0.98
3		**Gleaf2**	0.95	0.92	0.98
4		**Gleaf3**	0.79	0.92	0.93
5		**Gleaf4**	0.81	0.97	1.00
6		**Gleaf5**	0.99	0.99	0.99
7		**Groot**	0.92	0.93	0.93
8		**Gstem**	0.97	0.94	0.95
9		**Gstamen**	0.61	0.61	0.97
10		**Gpetal**	0.88	0.84	0.84
11		**Gcarpel**	0.99	0.61	0.57
12		**Gsepal**	1.00	0.97	0.97
13		**Gseedling2**	0.99	0.98	0.99
14		**Gseedling4**	0.90	0.92	0.99
15		**Gseedling6**	0.70	0.99	0.75
16		**Gseed1**	0.99	0.99	1.00
17		**Gseed2**	1.00	1.00	1.00
18		**Gseed3**	0.77	0.64	0.94
19	*T. hassleriana*	**Hleaf0**	0.97	0.97	0.99
20		**Hleaf1**	0.97	0.98	0.98
21		**Hleaf2**	0.96	0.98	0.98
22		**Hleaf3**	0.96	0.99	0.98
23		**Hleaf4**	0.96	0.99	0.98
24		**Hleaf5**	0.97	0.99	0.98
25		**Hroot**	0.95	0.96	0.96
26		**Hstem**	0.23	0.62	0.87
27		**Hstamen**	0.94	0.91	0.98
28		**Hpetal**	0.98	0.97	0.97
29		**Hcarpel**	0.95	0.99	0.98
30		**Hsepal**	0.87	0.86	0.90
31		**Hseedling2**	0.99	0.99	0.98
32		**Hseedling4**	0.99	1.00	0.99
33		**Hseedling6**	0.82	0.82	0.98
34		**Hseed1**	0.99	1.00	0.99
35		**Hseed2**	1.00	1.00	1.00
36		**Hseed3**	0.93	0.96	0.95

Supplemental Table 3 online. Pearson's correlation (r) of each individual replicate per tissue in G. gynandra and T. hassleriana respectively (A). Pearson's correlation between G. gynandra and T. hassleriana individual tissues (B).

B

Pearson Correlation *r* between G. *gynandra* and *T. hassleriana*		
#	**Tissue**	**r**
1	**Leaf0**	0.723369664
2	**Leaf1**	0.693967315
3	**Leaf2**	0.774414647
4	**Leaf3**	0.718280077
5	**Leaf4**	0.845767325
6	**Leaf5**	0.801946455
7	**Root**	0.693418487
8	**Stem**	0.397920288
9	**Stamen**	0.465027959
10	**Petal**	0.296842384
11	**Carpel**	0.409336161
12	**Sepal**	0.216833607
13	**Seedling2**	0.864093832
14	**Seedling4**	0.79602302
15	**Seedling6**	0.757896499
16	**Seed1**	0.922002838
17	**Seed2**	0.882400443
18	**Seed3**	0.612106172

Supplemental Table 4 online. Number of significatly up- or downregulated genes in *G. gynandra* compared to *T. hassleriana* within the different tissues. Differential expressed gene p-Values were calculated via EdgeR and Bonferroni-Holms corrected, genes with p<0.05 were classified as differential regulated.

Tissue	UP p< 0.05	UP p< 0.01	UP p< 0.001	DOWN p< 0.05	DOWN p< 0.01	DOWN p< 0.001
leaf0	5435	5061	4539	6076	5696	5237
leaf1	5197	4841	4391	5914	5529	5026
leaf2	4234	3894	3443	5047	4644	4204
leaf3	4646	4283	3833	5484	5070	4576
leaf4	3250	2911	2511	3774	3399	2979
leaf5	3236	2894	2447	4133	3716	3191
root	4343	3973	3511	5151	4755	4254
stem	7835	7497	7123	8462	8129	7698
stamen	4545	4116	3652	5388	4976	4451
petal	4445	4063	3613	5122	4751	4317
carpel	3718	3352	2929	3640	3274	2894
sepal	5650	5276	4780	6422	6023	5539
seedling2	4012	3644	3186	4354	3981	3546
seedling4	4113	3684	3202	4416	4043	3569
seedling6	2874	2534	2180	3542	3154	2714
seed1	4116	3764	3321	4457	4083	3591
seed2	6600	6270	5807	7075	6727	6276
seed3	6108	5725	5307	7088	6674	6190
mean	4686.5	4321.222222	3876.388889	5308.055556	4923.555556	4458.444444
max	7835	7497	7123	8462	8129	7698

Supplemental Table 5 online. List of genes present in root to shoot recruitment module.

T. hassleriana cds ID (Cheng et al., 2013)	Arabidopsis homologue	Coexpressed with TF	TAIR short annotation
T.hassleriana_10164	AT1G70410		beta carbonic anhydrase 4
T.hassleriana_20805	AT2G22500		uncoupling protein 5
T.hassleriana_17885	**AT5G61590**	**ERF**	**Integrase-type DNA-binding superfamily protein**
T.hassleriana_27615	**AT1G04250**	**Aux/IAA**	**AUX/IAA transcriptional regulator family protein**
T.hassleriana_13599	**AT5G13180**	**VND-I2**	**NAC domain containing protein 83**
T.hassleriana_07159	AT4G12730	Aux/IAA	FASCICLIN-like arabinogalactan 2
T.hassleriana_22160	AT5G57560		Xyloglucan endotransglucosylase/hydrolase family protein
T.hassleriana_03276	AT1G11545	Aux/IAA	xyloglucan endotransglucosylase/hydrolase 8
T.hassleriana_11774	AT1G43670		Inositol monophosphatase family protein
T.hassleriana_19959	AT5G19140	ERF	Aluminium induced protein with YGL and LRDR motifs
T.hassleriana_13658	AT1G25230	ERF	Calcineurin-like metallo-phosphoesterase superfamily protein
T.hassleriana_11758	AT3G14690	VND-I2	cytochrome P450, family 72, subfamily A, polypeptide 15
T.hassleriana_00726	AT5G46900		Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily
T.hassleriana_13312	AT3G22120		cell wall-plasma membrane linker protein
T.hassleriana_18867	AT3G54110		plant uncoupling mitochondrial protein 1
T.hassleriana_22110	AT1G14870		PLANT CADMIUM RESISTANCE 2
T.hassleriana_13333	AT5G19190		
T.hassleriana_11698	AT3G13950		
T.hassleriana_01980	AT5G25265		
T.hassleriana_04483	AT5G62900		
T.hassleriana_21987	AT1G13700	ERF	6-phosphogluconolactonase 1
T.hassleriana_15837	AT1G05000		Phosphotyrosine protein phosphatases superfamily protein
T.hassleriana_08797	AT5G23750	Aux/IAA	Remorin family protein
T.hassleriana_08517	AT5G36160		Tyrosine transaminase family protein
T.hassleriana_12936	AT5G25980		glucoside glucohydrolase 2
T.hassleriana_04639	AT2G01660		plasmodesmata-located protein 6
T.hassleriana_22812	AT4G21870	ERF	HSP20-like chaperones superfamily protein
T.hassleriana_10363	AT3G11660	VND-I2	NDR1/HIN1-like 1
T.hassleriana_19882	AT3G04720		pathogenesis-related 4
T.hassleriana_27070	AT2G15220		Plant basic secretory protein (BSP) family protein
T.hassleriana_05312	AT2G37170		plasma membrane intrinsic protein 2
T.hassleriana_05313	AT2G37170		plasma membrane intrinsic protein 2
T.hassleriana_12285	AT2G36830	Aux/IAA	gamma tonoplast intrinsic protein
T.hassleriana_12284	AT2G36830		gamma tonoplast intrinsic protein
T.hassleriana_14369	AT1G11670	Aux/IAA	MATE efflux family protein
T.hassleriana_08980	N.A.		
T.hassleriana_07000	N.A.		

Supplemental Table online 6. List of clustered general leaf developmental and vasculature regulating genes along both leaf gradients.

T. hassleriana cds ID (Cheng et al., 2013)	AGI	Annotation based on TAIR10	Function in vascular development
T.hassleriana_16883	AT1G19850	MONOPTEROS (MP)	leaf initiation
T.hassleriana_08823	AT1G19850	MONOPTEROS (MP)	leaf initiation
T.hassleriana_08424	AT1G32240	KANADI 2 (KAN2)	leaf axis formation
T.hassleriana_09176	AT1G32240	KANADI 2 (KAN2)	leaf axis formation
T.hassleriana_20498	AT1G52150	ATHB-15	neg reg of vasc cell diff
T.hassleriana_09793	AT1G52150	ATHB-15	neg reg of vasc cell diff
T.hassleriana_06450	AT1G65620	ASYMMETRIC LEAVES 2 (AS2)	leaf initiation
T.hassleriana_19648	AT1G73590	PIN-FORMED 1 (PIN1)	vein initiation (polar auxin transport)
T.hassleriana_01843	AT1G79430	ALTERED PHLOEM DEVELOPMENT (APL)	vascular cell identity repressed by REV
T.hassleriana_19440	AT1G79430	ALTERED PHLOEM DEVELOPMENT (APL)	vascular cell identity repressed by REV
T.hassleriana_27016	AT2G13820	Bifunctional inhibitor/lipid-transfer protein	vein formation (xylogen)
T.hassleriana_27989	AT2G27230	LONESOME HIGHWAY (LHW)	transcription factor-related
T.hassleriana_09087	AT2G27230	LONESOME HIGHWAY (LHW)	transcription factor-related
T.hassleriana_15265	AT2G27230	LONESOME HIGHWAY (LHW)	transcription factor-related
T.hassleriana_15152	AT2G28510	Dof-type zinc finger DNA-binding family protein	Dof-type zinc finger DNA-binding family protein
T.hassleriana_27908	AT2G28510	Dof-type zinc finger DNA-binding family protein	Dof-type zinc finger DNA-binding family protein
T.hassleriana_06822	AT2G33860	ETTIN (ETT)	leaf axis formation abaxial fate
T.hassleriana_23279	AT2G33860	ETTIN (ETT)	leaf axis formation abaxial fate
T.hassleriana_23086	AT2G37630	ASYMMETRIC LEAVES 1 (AS1)	leaf initiation
T.hassleriana_18733	AT4G08150	KNOTTED-like from Arabidopsis thaliana (KNAT1)	leaf initiation
T.hassleriana_09854	AT4G08150	KNOTTED-like from Arabidopsis thaliana (KNAT1)	leaf initiation
T.hassleriana_25576	AT4G24060	Dof-type zinc finger DNA-binding family protein	Dof-type zinc finger DNA-binding family protein
T.hassleriana_22410	AT4G32880	homeobox gene 8 (HB-8)	vein initiation (post auxin marker of vascular patterning)
T.hassleriana_28697	AT5G16560	KANADI (KAN)	leaf axis formation abaxial; neg reg of PIN1
T.hassleriana_19776	AT5G16560	KANADI (KAN)	leaf axis formation abaxial; neg reg of PIN1
T.hassleriana_18288	AT5G60200	TARGET OF MONOPTEROS 6 (TMO6)	TARGET OF MONOPTEROS 6
T.hassleriana_16642	AT5G60200	TARGET OF MONOPTEROS 6 (TMO6)	TARGET OF MONOPTEROS 6
T.hassleriana_18265	AT5G60690	REVOLUTA (REV)	adaxial leaf axis formation
T.hassleriana_19132	AT5G60690	REVOLUTA (REV)	adaxial leaf axis formation
T.hassleriana_17767	AT5G64080	XYP1	vein formation (xylogen)
T.hassleriana_26861	AT5G64080	XYP1	vein formation (xylogen)

Supplemental Methods

Leaf clearings and safranine staining (Supplemental Figure 1)

For leaf clearings *T. hassleriana* and *G. gynandra* leaves of stage 0 to 5 were destained in 70% EtOH with 1% glycerol added for 24 hrs and cleared in 5% NaOH until they appeared translucent and rinsed with H_2O_{dest}. Leaves were imaged under dark field settings with stereo microcope SMZ1500 (Nikon, Japan). Prior safranine staining, leaves were destained with increasing EtOH series until 100% EtOH and stained for 5 -10 min with 1% safranine (1g per 100ml 96% EtOH). After destaining leaves were analyzed with bright field microscope (Zeiss, Germany). Vein orders were determined by width and position as described by (McKown and Dengler, 2009) for Flaveria species.

Contig assembly and annotation (Supplemental Figure 4, Table 1 and Dataset 3)

Cleaned and filtered paired end (PE) reads were used to create a reference transcriptome for each species. The initial *de novo* assembly was optimized by using 31-kmer using Velvet (v1.2.07) and Oases (v0.2.08) pipeline (Zerbino and Birney, 2008; Schulz et al., 2012). For quality purposes the longest assembled transcript was selected with custom made perl scripts if multiple contigs were present (Schliesky et al., 2012) resulting in 59,471 *G. gynandra* and 52,479 *T. hassleriana* contigs. For quality assessment PE reads were aligned again to the respective contigs for each species via TopHat standard settings with over 60% backmapping efficiency in both species. Assembled longest transcripts were annotated using BLASTX mapping against TAIR10 proteome database (cut-off $1e^{-10}$). The best blastx hits were filtered by the highest bitscore. For quality assessment of contigs, *T. hassleriana* contigs were aligned with BLASTN against *T. hassleriana* predicted cds (Cheng et al., 2013). Multiple matching contigs to one cds identifier were filtered with customized perl script.

Cross species mapping sensitivity assessment (Supplemental Figure 5; Table 2)

All three biological replicates of leaf stage 5, stamen and young seed from *T. hassleriana* were mapped with BLAT V35 in dnax mode (nucleotide sequence of query and reference are translated in six frames to protein) with default parameters to both, the *T. hassleriana* gene models and the *A. thaliana* TAIR10 representative gene models. Subsequently, the BLAT output was filtered for the best match per read based on the highest score. RPKMs were calculated based on mappable reads per million (RPKM). The RPKM expression data was collapsed to single *A. thaliana* AGIs (RPKM were added) to avoid multiple assigned *T. hassleriana*‘s IDs to the same AGI. Pearson’s correlation r was calculated between the mapped *T. hassleriana* replicates mapped on *A. thaliana* gene models among each other. Also Pearson’s correlation r was calculated between cross species mapped *T. hassleriana* leaf5, stamen and seed1 replicates and the replicates of Leaf5 mapped to its own cds in *T. hassleriana*.

Principal component analysis (Supplemental Figure 8)

Principal component analyses (PCA,Yeung and Ruzzo, 2001) was carried out with MULTI EXPERIMENT VIEWER VERSION 4 (MEV4, (Saeed et al., 2003; Saeed et al.,

2006) on gene row SD normalized averaged RPKMs with median centering.

Enzyme Assays (Supplemental Figure 15)

From *G. gynandra* leaf stage 2 to 5, enzymatic activities of known C_4 enzymes were determinedas summarized by Ashton et al. (1990) in three biological replicates.

Comparison of Cleomaceae leaf gradients to *A. thaliana* leaf differentiation (Supplemental Figure 19)

Examination of Cleomaceae expression patterns of genes differentially regulated during the transition from cell proliferation to expansion in *A. thaliana*.
Andriankaja et al. (2012) observed that the transition between cell proliferation and expansion occurred between days 9 and 10. They defined two sets of genes significantly differentially expressed between day 9 and 10, one up-regulated and one down-regulated. The expression of the *T. hassleriana* and *G. gynandra* homologues of these genes were analyzed. The sum of standard error (SSE) was taken as a quality control to determine an appropriate number of clusters. The number of cluster centers chosen was 7 and 5 for up-regulated and down-regulated genes, respectively. The *K*-means clustering was performed the same as before, except that genes were not previously filtered by expression level and genes were only binned once into clusters.

Supplemental References

Andriankaja, M., Dhondt, S., De Bodt, S., Vanhaeren, H., Coppens, F., De Milde, L., Muehlenbock, P., Skirycz, A., Gonzalez, N., Beemster, G.T.S., and Inze, D. (2012). Exit from Proliferation during Leaf Development in Arabidopsis thaliana: A Not-So-Gradual Process. Dev. Cell 22, 64-78.

Ashton A.R., Burnell J.N., Furbank R.T., Jenkins C.L.D., Hatch M.D. (1990). The enzymes in C4 photosynthesis. In Enzymes of Primary Metabolism. Methods in Plant Biochemistry, P.M. Dey and J.B. Harborne, eds (London: Academic Press), pp. 39–72

Cheng, S., van den Bergh, E., Zeng, P., Zhong, X., Xu, J., Liu, X., Hofberger, J., de Bruijn, S., Bhide, A.S., Kuelahoglu, C., Bian, C., Chen, J., Fan, G., Kaufmann, K., Hall, J.C., Becker, A., Braeutigam, A., Weber, A.P.M., Shi, C., Zheng, Z., Li, W., Lv, M., Tao, Y., Wang, J., Zou, H., Quan, Z., Hibberd, J.M., Zhang, G., Zhu, X.-G., Xu, X., and Schranz, M.E. (2013). The T arenaya hassleriana Genome Provides Insight into Reproductive Trait and Genome Evolution of Crucifers. Plant Cell **25,** 2813-2830.

McKown, A.D., and Dengler, N.G. (2009). Shifts in leaf vein density through accelerated vein formation in C-4 Flaveria (Asteraceae). Annals of Botany **104,** 1085-

1098.

Saeed, A.I., Hagabati, N.K., Braisted, J.C., Liang, W., Sharov, V., Howe, E.A., Li, J., Thiagarajan, M., White, J.A., and Quackenbush, J. (2006). TM4 microarray software suite. In DNA Microarrays, Part B: Databases and Statistics, A. Kimmel and B. Oluver, eds, pp. 134.

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.

Schliesky, S., Gowik, U., Weber, A.P.M., and Brautigam, A. (2012). RNA-Seq Assembly - Are We There Yet? Front Plant Sci **3,** 220-220.

Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA- seq assembly across the dynamic range of expression levels. Bioinformatics **28,** 1086- 1092.

Comparative Transcriptome Atlases Reveal Altered Gene Expression Modules between Two Cleomaceae C_3 and C_4 Plant Species

Canan Külahoglu, Alisandra K. Denton, Manuel Sommer, Janina Maß, Simon Schliesky, Thomas J. Wrobel, Barbara Berckmans, Elsa Gongora-Castillo, C. Robin Buell, Rüdiger Simon, Lieven De Veylder, Andrea Bräutigam and Andreas P.M. Weber

Plant Cell; originally published online August 8, 2014;
DOI 10.1105/tpc.114.123752

This information is current as of September 18, 2014

Supplemental Data	**http://www.plantcell.org/content/suppl/2014/07/09/tpc.114.123752.DC1.html**
Permissions	https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&issn=1532298X&WT.mc_id=pd_hw1532298X
eTOCs	Sign up for eTOCs at: http://www.plantcell.org/cgi/alerts/ctmain
CiteTrack Alerts	Sign up for CiteTrack Alerts at: http://www.plantcell.org/cgi/alerts/ctmain
Subscription Information	Subscription Information for *The Plant Cell* and *Plant Physiology* is available at: http://www.aspb.org/publications/subscriptions.cfm

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY

Manuscript-draft 3

Glk1 function is conserved between *Cleome gynandra* and *Arabidopsis thaliana* as indicated by partial rescue of the broadly de-regulated transcriptome of the *glk*1 *glk*2 double mutant

Manuel Sommer*, Andrea Bräutigam, Andreas PM Weber[a]

Plant Biochemistry, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany

* first author

[a] corresponding author

Abstract

The Golden2-like (Glk) genes are the first and only known regulators of photosynthesis and chloroplast development in higher plants. A pair of Glk genes acts redundantly in the C_3 plant *Arabidopsis thaliana*, while monocotyledonous C_4 plants compartmentalize Glk gene expression between mesophyll (M) and bundle sheath (BS) to drive tissue-specific plastid development. In contrast, only a single Glk is expressed in both M and BS of the dicotyledonous C_4 plant *Cleome gynandra* (*Cg*Glk1). In this study, we tested functional redundancy between the ubiquitously expressed Glk genes from *A. thaliana* and functional conservation of Glk1 function between *A. thaliana* and *C. gynandra* on a molecular level. Therefore, we compared the transcriptome of *A. thaliana* wildtype with *glk*1 and *glk*2 single mutants, the *glk*1 *glk*2 double mutant and two lines that express *Cg*Glk1. We monitored the expression of 22882 transcripts and filtered the dataset by a threshold of 20 reads per kilobase and million. Of 5433 remaining genes, 1145 were differentially expressed between wildtype and *glk1 glk2*. This high number of de-regulated genes exceeded the known targets of Glk genes by far and involved most photosynthetic genes. Also groups of non-photosynthetic proteins showed up to 33% of differentially expressed genes. Surprisingly, we could show that transcripts of calvin cycle enzymes remain unaffected by the widespread changes. The transcriptome of the *glk*1 single mutant was very similar to the wildtype, but the *glk*2 mutant had 110 genes differentially expressed, 100 of which were down-regulated. The group of down-regulated genes contained ten photosynthesis and tetrapyrrole biosynthesis genes that represent novel targets of Glk2. Light expression of *Cg*Glk1 in the double mutant led to significant complementation in 376 genes, leaving 67% of all affected genes un-rescued. Photosynthetic genes were most frequently rescued, which indicates their conservation as primary targets of Glk genes between *A. thaliana* and *C. gynandra*. Accordingly, expression of *Cg*Glk1 also rescued the pale green phenotype of the mutant. The relative degree, to which the mRNA levels were rescued, was significantly lower for PSII core genes and especially those that are part of the oxygen evolving complex (OEC) (37%) than for genes of the light harvesting complex (LHC) of PSII (95%), which might be a result of a changed target specificity of *Cg*Glk1 as compared to *At*Glk1. We further tested the effect of Glk mutation and complementation on the photosynthetic apparatus by physiological analysis. The *glk*1 *glk*2 double mutant showed a decrease in photosynthetic performance at both moderate light (100 µE) and high-light (1500 µE) and displayed impaired photoprotection. Mild expression of *Cg*Glk1 rescued the photosynthetic phenotype and the quantum yield of PSII partially, but was not able to rescue the high amount of NPQ. The *glk*2 single mutant behaved wildtype-like despite the changes on transcript level. Our results suggest that the targets of Glk1 and Glk2 are not fully redundant in leaves of *A. thaliana* but loss of a single gene does not cause a physiologically detectable phenotype. The double mutant, however, causes de-regulation of numerous photosynthetic and non-photosynthetic genes. We also

conclude that Glk1 function is largely conserved between *A. thaliana* and *C. gynandra*, as indicated by the rescued visual phenotype and the reduced impairment of gene expression, photosynthesis and photoprotection. To determine, whether the low complementation of OEC genes by *Cg*Glk1 results from changed target specificity, we will screen for complementing lines that show very high *Cg*Glk1 expression and quantify signature genes of the OEC and the LHC by qRT-PCR in future experiments.

Introduction

Photoautotrophic organisms use CO_2 and H_2O to produce hexose sugars and O_2. The central carboxylating enzyme in photosynthesis, ribulose-1,5-bisphosphate oxygenase/carboxylase (RuBisCO), evolved 2.7 billion years ago (Buick, 1992) in an atmosphere that rarely contained molecular O_2 (Farquhar et al., 2011). This lack of selective pressure prevented the enzyme from evolving high substrate specificity to distinguish between the chemically very similar gases O_2 and CO_2. To avoid a loss in biomass-gain through this oxygenation, plants evolved carbon concentrating mechanisms (CCMs) around RuBisCO (Raven et al., 2008).

C_4 plants concentrate carbon by compartmentalizing photosynthetic processes between two leaf tissues and strongly enriching CO_2 over O_2 in one tissue. This compartmentalization implicates the need of specialized gene expression profiles for each tissue. CO_2 enrichment occurs in the bundle sheath (BS) cell layer in the leaf center. It harbors the vast majority of the plants RuBisCO and shows enrichment in photosystem I (PSI), while the outer mesophyll (M) cell layer contains a CO_2-prefixing enzyme and photosystem II (PSII)-rich thylakoids. Low expression of PSII in the BS is essential to minimize oxygen evolution around RuBisCO. For spatial CO_2 enrichment, a metabolic cycle transports prefixed CO_2 from M to BS tissue to overcome the RuBisCO oxygenation reaction. This shuttling pathway and numerous secondary adaptations like reallocation of redox reactions to the reducing M environment make C_4 photosynthesis a very complex trait (Moore and Black, 1979; Majeran et al., 2005; Majeran and van Wijk, 2009).

The complexity of C_4 metabolism is in contrast to scarce knowledge about its transcriptional regulation and organization with only one exception, namely the *Golden2-like* (*Glk*) transcription factors (Waters et al., 2009; Gowik and Westhoff, 2011). *Glk* genes were shown to regulate chloroplast development in C_3 and C_4 plants, whereat the evolution of C_4 photosynthesis is always preceded by Glk gene duplication of the single, ancestral Glk gene (Rossini et al., 2001; Waters et al., 2009; Wang et al., 2013). When two copies of Glk exist, the leaf expression pattern of *Glk1* and *Glk2* is redundant in C_3 plants and compartmentalized between BS and M in the monocotyledonous C_4 plants sorghum and maize (Wang et al., 2013). Microarray analysis of *glk1 glk2* double mutants of *A. thaliana* upon induction of Glk expression demonstrated that the *Glk* genes mainly regulate photosynthetic genes in the C_3 background in a redundant manner. The set of detected Glk targets contained 66 genes and

was dominated by genes involved in chlorophyll-biosynthesis and formation of PSII (Waters et al., 2009). However, transcriptome data of *glk1 glk2* were not yet compared to a wildtype transcriptome and thus, the full impact of a loss of these transcription factors on other photosynthetic and non-photosynthetic genes remains to be determined. The pale green phenotype of *glk1 glk2* can be complemented by expression of the Glk1 gene of the moss *Physcomitrella patens*, showing conservation of Glk gene function between two very distant C_3 land plants (Yasumura et al., 2005). It is thus broadly accepted that Glk function is conserved throughout the plant kingdom. Conservation of Glk function between a C_3 and a C_4 plant, however, was not yet shown. In maize, expression of the *Golden2* gene is predominantly located to undifferentiated leaf tissue and BS cells, which causes the *golden*2 mutant to display impaired chloroplast development mainly in BS cells (Hall et al., 1998; Rossini et al., 2001). This phenotype suggests a key function of Glk genes in the regulation of photosynthetic gene expression in monocotyledonous C_4 plants, too. The dicotyledonous C_4 plant *C. gynandra* is the closest related C_4 species to *A. thaliana* (Brown et al., 2005). It shows no compartmentalization of Glk genes, but rather expresses only Glk1 in both M and BS cells whereas Glk2 is rarely expressed (Wang et al., 2013). Glk mediated, cell-type specific chloroplast development is thus no universal characteristic of C_4 photosynthesis. The non-compartmentalized expression pattern of *Cg*Glk1 and its exclusive expression make it an ideal target to analyze functional conservation between Glk genes of C_3 and C_4 plants. Assuming that Glk function is conserved, heterologous expression of *Cg*Glk1 is expected to cause full complementation of the *A. thaliana glk1 glk2* phenotype. However, this hypothesis is challenged by the fact that in maize, where Glk genes are expressed in a tissue-specific manner, both cell types show very distinct expression patterns of photosynthetic genes (Majeran et al., 2008). This raised hypotheses about cell-type specific functional adaptation of Glk genes to C_4 (Rossini et al., 2001; Wang et al., 2013). It was also shown, that Glk specializes during evolution, as the *P. patens glk1 glk2* mutant was not complemented by a functional copy of the more derived *At*Glk1 (Bravo-Garcia et al., 2009).

In this study, we tested functional conservation between the ubiquitously expressed *Cg*Glk1 and the redundant Glk genes in the C_3 plant *A. thaliana* by heterologous expression of *Cg*Glk1 in the *glk1 glk2* mutant. We monitored global changes in regulation of transcription by RNA sequencing and complemented our findings in physiological studies.

Material and Methods

Chemicals and reagents

Chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Reagents for the ß-Glucuronidase assay were purchased from Duchefa (Haarlem, Netherlands)

Generation of plant lines and cultivation

Glk mutants of *A.thaliana* were ordered from the European Arabidopsis Stock Center (uNASC). Stock numbers were N9805 and N9806 for the single mutants, *glk1* and *glk2*, respectively. The stock number of the *glk1 glk2* double mutant was N9807. To generate the complementing lines, the *Cg*Glk1 gene sequence (Bräutigam 2011, amplified via PCR using primers P1/P2) was cloned into pCambia1381 (Cambia, Canberra, Australia) over SpeI/BstEII restriction sites. In a second step, the native *Cg*Glk1 promoter (amplified via PCR using primers P3/P4) was fused to the gene over BamHI/SpeI restriction sites for co1. In a similar approach, the 35S promoter from pCambia1301 (amplified via PCR using P5/P6) was fused to the gene over BamHI/SpeI restriction sites for co2. *Agrobacterium tumefaciens* strain GV2260 (Deblaere et al., 1985) was transformed with the resulting plasmid and flowering *A. thaliana* plants were infected with the bacteria by floral dip (Clough 1998). The offspring of infected plants was screened on solid MS media containing Hygromycin (33 µg/mL). 17 and 16 plants that were transformed with the co1 and co2 construct, respectively, grew on selective plates. The plants were selfed, screened for phenotypical complementation and *Cg*Glk1 expression was monitored in complemented individuals (Supplemental Figures 1, 2 and 3).

For cultivation, *A. thaliana* plants were vernalized and subsequently germinated in petri dishes on solid MS media in a Percival CU-36L5/D incubator (Percival, Perry, IA). Twelve days after germination, plants were transferred to 9 cm pots containing soil and were grown in Percival CU-36L5 incubators (Percival, Perry, IA). In both incubators, plants were illuminated 12 hours a day with a light intensity of 100 µE at 22°C during the day and 18°C during the night. Plants were watered with 0.2% Wuxal 8+8+6 fertilizer (Bayer, Leverkusen, Germany) every seven days. Twenty-eight days after germination, plants were used for experiments. When leaf extract was required for an experiment, three leaves of three plants were cut and frozen in liquid nitrogen immediately (< 1s). Frozen leaves were then ground to a fine powder under constant supply with liquid nitrogen using mortar and pestle. The powder was stored at -80°C until used.

Phylogenetic analysis of Glk genes

Protein sequences of maize Glks were extracted from the non-redundant NCBI Protein database (www.ncbi.nlm.nih.gov/protein). *C. gynandra* and *Tarenaia hassleriana* sequences were translated from genome and transcriptome information (Cheng et al., 2013; Külahoglu et al., 2014). *A. thaliana* sequences were extracted from TAIR (www.arabidopsis.org). Phylogenetic analysis was carried out on phylogeny.fr (www.phylogeny.lirmm.fr) using the MUSCLE alignment algorithm, Gblocks curation and the PhyML phylogeny software. Multiple sequence alignment was done with ClustalW2 (www.ebi.ac.uk/Tools/msa/clustalw2) and edited in Microsoft Word (Microsoft, Redmond, WA).

Next-Generation sequencing of leaf mRNA of nuclear encoded genes

Leaf extract of 28 day old plants was used for RNA-seq sample preparation. Total RNA was extracted with the RNeasy Mini Kit (Qiagen, Hilden, Germany), DNase treated and quality controlled on a Bioanalyzer (2100, Agilent, Santa Clara, CA; data not shown). Subsequently, mRNA purification and adapter ligation was done with the TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA) using 1 µg of total RNA. After a second quality control on the bioanalyzer (data not shown), samples were subjected to Beckmann Genomics (Danvers, MA) and sequenced on a HiSeq2500 sequencer (Illumina, San Diego, CA) for single end sequencing of >169 million 100-basepair reads according to the manufacturer's instructions.

Analysis of leaf transcriptomes

The transcriptomes were mapped to the *A. thaliana* TAIR 9 coding sequence database using the CLC Workbench (CLC Bio, Aarhus, Denmark). Similarity analysis was performed with R (R: A Language and Environment for Statistical Computing, Vienna, Austria) using the EdgeR toolkit for PCA (Robinson et al., 2010) and via Hierarchical Clustering (Pearson Correlation, Average Linkage) using the Multi-Experiment-Viewer (MEV, Dana-Farber Cancer Institute, Boston, MA). Expression data were transformed into mean-centered Z-scores before analysis with MEV. Differential expression was also tested with EdgeR, applying Fisher's Exact Test and multiple hypothesis correction after Benjamini-Hochberg, based on the assumption of a negative binomial distribution of data. Over-representation analysis was performed with the PageMan tool, integrated into the MapMan software (Usadel et al., 2006). Remaining calculations and enrichment analyses were done with Excel (Microsoft, Redmond, WA) and visualized with Prism (GraphPad, LaJolla, CA). MapMan (Thimm et al., 2004) bins were used as a basis for enrichment analyses. The oxygen evolving complex (OEC) bin was designed by adding all genes that contain the term "oxygen evolving" in their TAIR 9 annotation. For analysis of signature genes, we excluded all genes with expression values below 1000 RPKM in the wildtype. The remaining six genes in the OEC bin were manually double-checked versus the MapMan database and published literature (Ferreira et al., 2004; Suorsa et al., 2006).

Chlorophyll determination

For the determination of total chlorophyll and chlorophyll A/B ratio, 50 mg of leaf extract were solved in 800 µL of phosphate buffered acetone, diluted with 200 µL of water and centrifuged 5 minutes. The supernatant was analyzed with a spectral photometer and chlorophyll concentrations were calculated according to Porra (Porra, 2002).

Expression of ß – glucuronidase in reporter lines and staining assay

Gateway cloning (Invitrogen, Carlsbad, CA) was used to integrate the *Cg*Glk1 promoter sequence (amplified via PCR with P1/P2) into pMDC163 (Curtis and Grossniklaus, 2003). *Agrobacterium tumefaciens* strain GV2260 (Deblaere et al., 1985) was transformed with the resulting plasmid and flowering *A. thaliana* plants were infected with the bacteria by floral dip (Clough and Bent, 1998). The offspring of infected plants was screened on solid MS media containing Hygromycin (33 µg/mL). Leaf disks of resistant plants were analyzed by a ß – Glucuronidase assay (Vitha, 2007).

Expression of a green fluorescent protein in reporter lines and microscopy

The *Cg*Glk1 gene (amplified via PCR with P3/P4) was integrated into pMDC83 (Curtis and Grossniklaus, 2003) using Gateway cloning. *Agrobacterium tumefaciens* strain GV2260 (Deblaere et al., 1985) was transformed with the resulting plasmid and *Nicotiana benthamiana* leaves were infiltrated with a suspension of bacterial cells (Romeis et al., 2001). Two days after infiltration, protoplasts were extracted from infected leaf tissue (Sheen, 1991) and observed under a laser scanning microscope (Zeiss LSM510, Carl Zeiss, Jena, Germany) with a GFP filter (excitation 488 nm; emission 509 nm). Chloroplast fluorescence was monitored with a DsRed filter (excitation 556 nm; emission 630 nm) and a brightfield picture was taken to visualize membranes. A digital merge of the pictures was assembled with the Zeiss Zen 2011 software (Carl Zeiss, Jena, Germany).

Quantification of metabolites

For metabolite quantification, leaf extract was solved in a 2.5:1:1 mixture of methanol:chloroform:H_2O and processed according to a standard protocol (Fiehn, 2007) and GC/MS (Agilent 7200 QTOF GC/MS, Agilent, Santa Clara, CA) was carried out according to Lee and Fiehn (Lee and Fiehn, 2008). Quantification of chromatograms was done with the MassHunter Workstation (Agilent, Santa Clara, CA).

Determination of leaf photosynthesis rates

Photosynthesis rates were determined using a LI-6400XT (LI-COR, Lincoln, NE). Leaves of 28 day old plants were clipped to the measuring device and light curves were monitored with three minutes incubation before each measurement. Measurements were done during the day, no closer than two hours to either dawn or dusk. Five or more replicates grown in two generations were measured for each genotype. Data were normalized to the average photosynthesis rate of wildtype plants at 2000 µE (7.64 and 9.56 µmol CO_2 m^{-2} s^{-1} in the first and second generation, respectively).

Chlorophyll fluorescence measurements

The maximal photochemical quantum yield of PSII was determined by measuring Fv/Fm values on leaf disks of 28 day old plants with a pulse modulated fluorometer (JUNIOR-PAM, Heinz Walz, Effeltrich, Germany). Leaf disks of dark adapted plants were placed on a cooled water bath and incubated with high light (1500 µE) for 100 minutes. The relative amount of non-photochemical quenching (NPQ) was also determined with a pulse modulated fluorometer (DUAL-PAM-100, Heinz Walz, Effeltrich, Germany). The measuring device was clipped to leaves and NPQ was monitored for 10 minutes at illumination with 825 µE and subsequently for 10 minutes in darkness.

Results

Cleomaceae contain two Glk genes, which are orthologs of *A. thaliana* Glk1 and Glk2

For our study, we extracted the coding sequence of *Cg*Glk1 and *Cg*Glk2 from previously published RNA sequencing data (Külahoglu et al., 2014). We analyzed their phylogenetic proximity to other Glk genes in the Brassicales (Cheng et al., 2013) and to the maize Glk genes based on the amino-acid sequence (Figure 1). We found that the Glk1 proteins of plants within the Brassicales are phylogenetically closer than the Glk1 protein and the Glk2 protein of one species. The maize genes form an outgroup, which is slightly closer to the Glk1 subtree. A multiple sequence alignment of Glk1 proteins (Figure 2) demonstrated high overall similarity. The conserved DNA binding domain was identical between *C. gynandra* and its close C_3 relative *T. hassleriana* and a single amino acid exchange occurred in the N-terminal helix sequence in *A. thaliana*. In the GCT box, the second conserved domain, eight amino acid exchanges occurred between *A. thaliana* and *C. gynandra*, while *T. hassleriana* Glk1 has lost large parts of this domain.

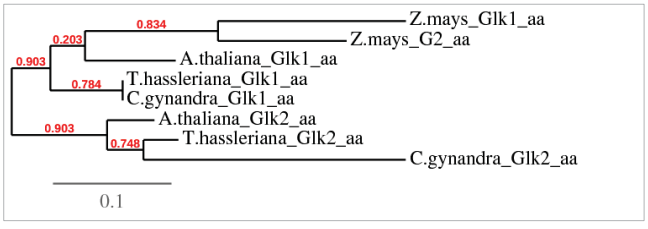


Figure 1: Phylogenetic tree of Glk protein sequences. Branch length is proportional to the number of substitutions per site. Red numbers indicate branch support.

Figure 2: Multiple sequence alignment of three Glk1 orthologs of the *Brassicales*. Underlines indicate the DNA-binding helix-loop-helix domain (three continuous lines) and the C-terminal GCT-box (dashed line).

*Cg*Glk1 complements the pale, dwarfish phenotype of the *glk1 glk2* mutant in *A. thaliana*

The *A. thaliana glk1 glk2* double mutant is pale green and retarded in growth, while the single glk mutants *glk1* and *glk2* are visually indistinguishable from the wildtype (Figure 3) (Fitter et al., 2002). We analyzed the potential of *Cg*Glk1 to complement the mutant phenotype via agrobacteria-mediated transformation of flowering double mutant plants with *Cg*Glk1 driven by either the *Cg*Glk1-promoter (line co1.1) or a 35S-promoter (line co2.1). The pale green, dwarfish phenotype of the *glk1 glk2* mutant was weakly rescued in co1.1, while full complementation of the visual phenotype was achieved in co2.1 (Figure 3). To quantify the complementing effect, we measured total chlorophyll levels and the chlorophyll a/b ratio of wildtype, mutants and complemented plants. The *glk1 glk2* mutant possessed 48% less chlorophyll and displayed a significantly higher Chlorophyll A/B ratio of 4.84, which describes a 37% increase versus the wildtype (Figure 4). The single mutants, in contrast to their uniform appearance, contain different amounts of chlorophyll. Both chlorophyll content and chlorophyll A/B ratio of *glk1* resemble wildtype values, whereas *glk2* displays mild similarity to the double mutant (14% chlorophyll reduction, 10% increased ratio). Restoration of chlorophyll parameters in the complemented lines varies depending on the promoter that drives CgGlk1. In co2.1, chlorophyll content and composition are as in *glk2*. In contrast, co1.1 plants possess 22% less chlorophyll than the wildtype and show an increased Chlorophyll A/B ratio by 22%.

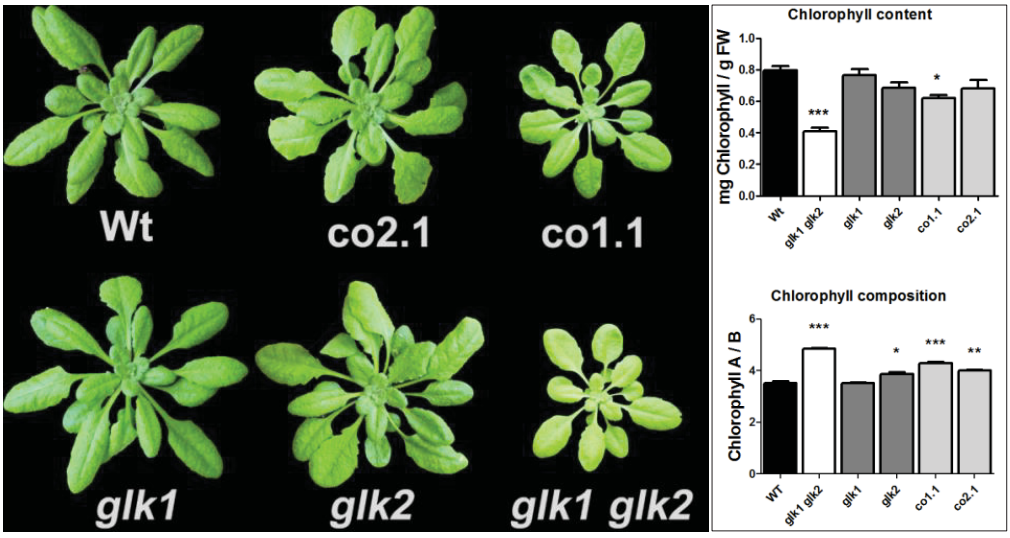


Figure 3: Leaf rosettes of 28 day old *A. thaliana* plants of wildtype, *glk* single and double mutant and two complementing lines, co1.1 and co2.1. Plants were grown on soil at 100 µE and 12 hours illumination/day.

Figure 4: Chlorophyll content of all genotypes was measured photometrically on an 80% acetone extract of leaf powder. Total chlorophyll (A) and chlorophyll a/b ratio (B) were derived from absorption values at 646, 663 and 750 nm.

*Cg*Glk1 is located to the nuclei of green leaf tissue when expressed in *A. thaliana*

In localization experiments with fusions to ß-Glucuronidase (GUS) and a green fluorescent protein (GFP), we tested whether *Cg*Glk1 is expressed in green tissues of *A.thaliana* and localizes to the nucleus, so it can potentially function as a photosynthetic transcription factor. The GUS reporter was detected uniformly spread across a leaf disk (Figure 5) and the GFP reporter was found exclusively in the nucleus of a protoplast suspension (Figure 6).

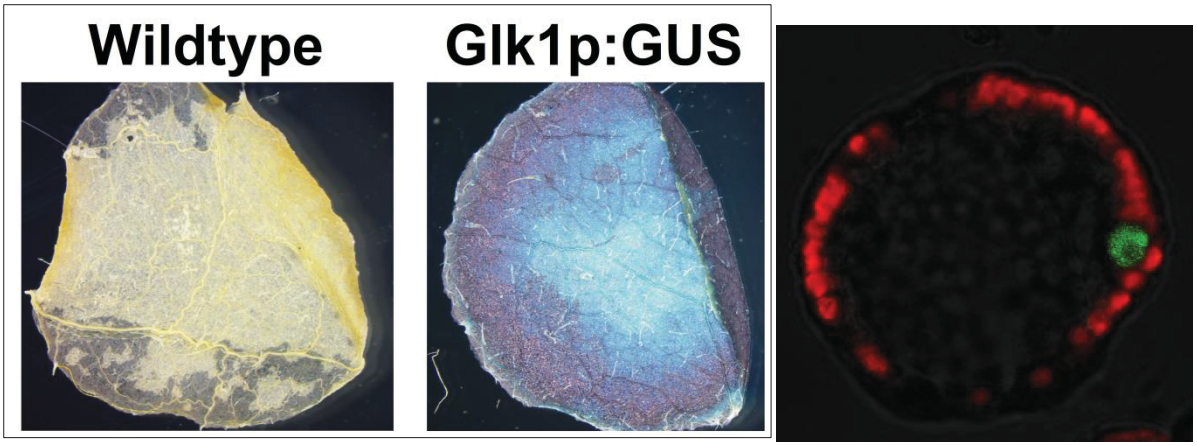


Figure 5: Leaf disks of 28 day old *A. thaliana* wildtype plants (left) and plants carrying a *Cg*Glk1promoter:GUS fusion (right) after a ß-Glucuronidase assay

Figure 6: *N. benthamiana* protoplast after transient transformation with *A. tumefaciens* harboring a *Cg*Glk1:GFP fusion construct. Picture is a merge of GFP fluorescence (green), chlorophyll autofluorescence (red) and a brightfield image (grayscale).

The transcriptomes of *glk* mutants and wildtype display characteristic differences

To reveal all genes that are affected by a lack of Glk expression and assess the potential of *Cg*Glk1 to rescue them on the transcript level, we conducted RNA sequencing of mRNA from leaves of wildtype, mutants and complemented plants that were grown in moderate light (100 µE). Single end sequencing in a HiSeq 2500 device yielded 196,312,162 raw reads. Of these, 185,330,247 reads (94%) mapped to the *A. thaliana* TAIR 9 coding sequence database, matching with a total of 22882 genes. The number of reads that were assigned to one sample ranged from 6.3M (in sample *glk1glk2*_1) to 14.1M (in sample co1.1_2, Table 1).

Differential gene expression between the genotypes was determined in R using the edgeR package and corrected for multiple hypothesis testing. Wildtype and *glk1 glk2* showed the greatest diversity between each other by having 8.8% of all genes differentially expressed (Table 2). The groups of genes that were up- and down-regulated in the mutant were of similar size (4.3% and 4.5%,

respectively). There were fewer genes differentially expressed in the single mutants, namely 0.1% in *glk1* and 0.5% in *glk2*. Intermediate numbers of genes were changed in the complemented lines with co1.1 having 6.3% of all genes differentially expressed and co2.1 3.6%. When the complemented lines were tested for differential gene expression against *glk1 glk2*, co1.1 showed higher similarity to the double mutant than to the wildtype (0.7% of genes different). In co2.1, we found 2.7% of all genes differentially expressed when compared to the double mutant.

Table 1: Number of reads generated by RNA sequencing per sample and respective numbers of reads that were uniquely mapped to the A. thaliana TAIR 9 database.

	total reads	unique reads
Wt_1	7554879	6656636
Wt_2	10553001	9905225
Wt_3	9174893	8575823
glk1glk2 _1	6284454	5732747
glk1glk2 _2	9963088	9350689
glk1glk2 _3	9742794	9210626
glk1 _1	11803874	11169851
glk1 _2	11171907	10651867
glk1 _3	11232596	10603176
glk2 _1	9334734	8904741
glk2 _2	12124875	11548825
glk2 _3	11841340	11267833
co1.1_1	10522191	9931726
co1.1_2	14067845	13395585
co1.1_3	11644046	11104517
co2.1_1	13179454	12569851
co2.1_2	13412839	12745740
co2.1_3	12703352	12004789
Σ	196312162	185330247

Table 2: Overview of the pairwise comparison of genotypes with EdgeR. Source for this overview is the differential expression table.

	Glk1glk2 vs. Wt	*glk1 vs. WT*	*glk2* vs. Wt	co1.1 vs. Wt	co2.1 vs. Wt	co1.1 vs. *glk1glk2*	co2.1 vs. *glk1glk2*
Up	993	7	10	663	367	143	366
Unchanged	20859	22865	22772	21442	22140	22713	22256
Down	1030	10	100	777	375	26	260
% Up	4.3	0	0	2.9	1.6	0.6	1.6
% Unchanged	91.2	99.9	99.5	93.7	96.8	99.3	97.3
% Down	4.5	0	0.4	3.4	1.6	0.1	1.1

Qualitative analysis of differentially expressed genes

We performed Principal component analysis to visualize variation between all replicates before we analyzed the samples qualitatively. The first two components explained 22% and 9% of variation, respectively (Figure 7). While the first principal component separates the samples by genotype, no such statement can be made for the second component based on our knowledge. Wildtype and single knockout samples partially overlap in the first component as well. Hierarchical clustering by Pearson correlation of the samples also did not separate wildtype samples from glk1 samples (Figure 8). To verify the complemented lines, we analyzed the transcript abundance of *Cg*Glk1. We counted 13.5 reads per kilobase and million (RPKM) in co1.1 and 38.0 RPKM in co2.1. Both values are significantly lower than *At*Glk expression in the wildtype (Figure 9).

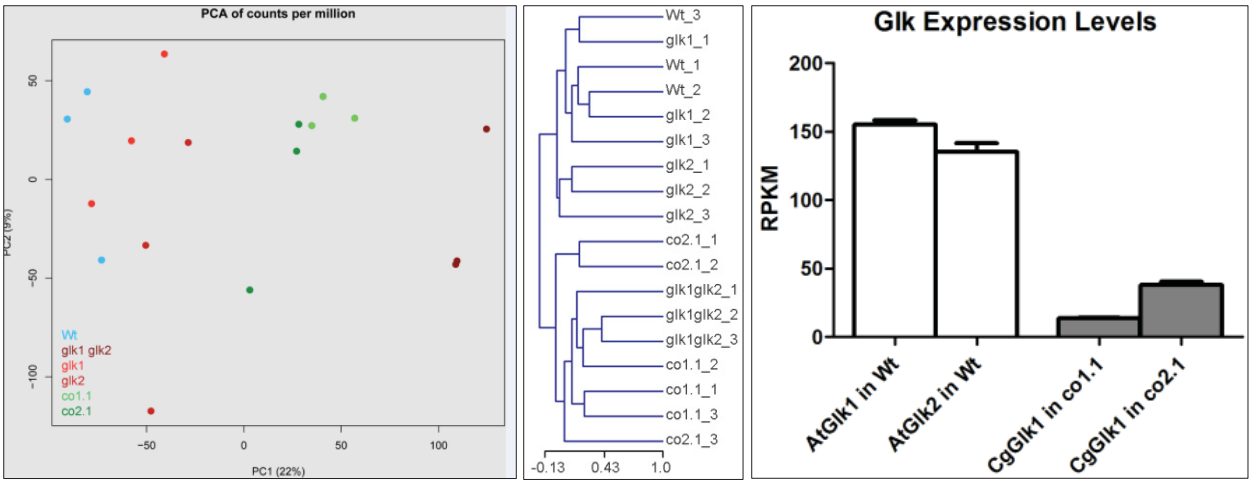


Figure 7: PCA of individual replicates. PCA was done on expression values normalized to reads per million with the R software tool.

Figure 8: Hierarchical clustering of individual replicates. Clustering was done using Pearson correlation and average linkage with the clustering algorithm.

Figure 9: Expression of *At*Glk genes in wildtype and expression of *Cg*Glk1 in complemented mutants. Expression is normalized to reads per million reads and kilobase gene length (RPKM).

Next we assessed the similarity of the set of differentially expressed genes of the six genotypes. We used a significance cut-off value of 20 rpkm for this analysis and excluded the single knockout lines, since they had very few differentially expressed genes left (9 and 68, of which 8 and 65 are also different in *glk1 glk2*). Of the remaining 488 down-regulated genes in the double mutant, 203 were not significantly down in at least one complementing line (Figure 10). The group of 391 genes that were down-regulated in at least one complementing line shared 336 genes with the double mutant. Of 657 up-regulated genes in the double mutant, 543 were no longer up in at least one complementing line (Figure 10). Of 503 up-regulated genes in one or more complementing line, 369 were in the *glk1 glk2* group, too. When we compared the significantly down-regulated genes from *glk*2 and *glk*1 *glk*2, we found that they overlapped almost completely. Only two down-regulated genes in *glk*2 were not down-regulated in *glk*1 *glk*2 (Figure 10).

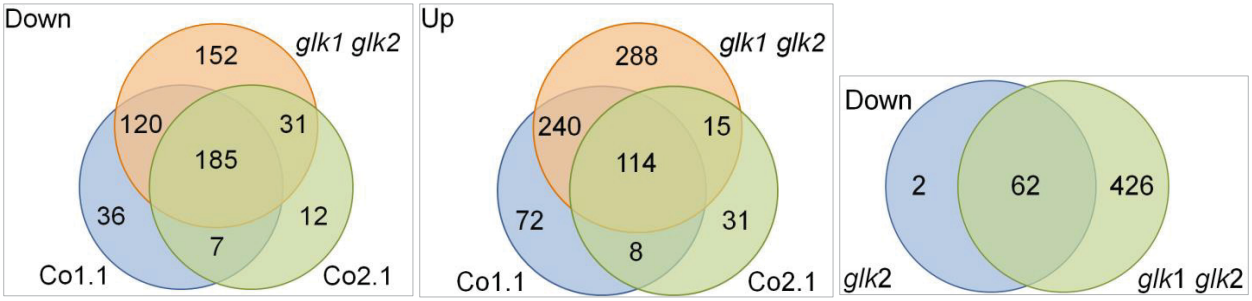


Figure 10: Venn diagrams of differentially expressed genes as compared to the wildtype. Classification into “Down” and “Up” followed expression differences between *glk1glk2* and wildtype. Genes that were only differentially expressed in a

complementing line were also classified according to the gene's expression difference between *glk1glk2* and wildtype, even though this was not significant. For *glk*2, a total of 4 up-regulated genes are not displayed as a Venn diagram.

To increase the quality of our prediction on rescued genes, however, we did not specify a gene as rescued when it was no longer differentially expressed between a complemented line and the wildtype. Instead it had to be differentially expressed between at least one complemented line and the *glk1 glk2* mutant. In this comparison, 376 genes matched this criteria and were also differentially expressed between wildtype and *glk1 glk2* (169 up-regulated and 207 down-regulated). This set of genes was referred to as complemented in further analyses.

Enrichment analysis of genes that are differentially expressed in *glk* mutants

When we categorized the 1145 differentially expressed genes of the *glk1 glk2* mutant to functional classes, genes involved in photosynthesis and chlorophyll biosynthesis appeared most frequently down-regulated (Figure 11). The percentage of down-regulated genes peaked in functional classes that contained genes of core complexes or LHCs of photosystems. This effect was observed in PSI and PSII. Genes were also frequently down-regulated in sulfur-containing secondary metabolism, serine-glycine-cysteine metabolism and other bins of primary and secondary metabolism. Functional groups with a high relative amount of up-regulated genes were N-metabolism, photorespiration, light signaling and DNA metabolism. As a complementary approach, we analyzed over- or under-representation of genes in all MapMan bins using PageMan (Figure 12). Among the down-regulated genes in *glk1 glk2*, photosynthetic lightreactions, secondary metabolism and auxin metabolism were over-represented, while DNA metabolism and protein synthesis were under-represented. The upregulated genes were over-represented by very few classes of mainly un-assigned proteins. Next, we visualized the differential expression of individual photosynthetic genes in the photosynthesis pathway of the MapMan tool (Figure 13). We found systematic downregulation of genes in glk1 glk2 that involved most genes of both PSI and PSII including electron carriers like plastocyanins and ferredoxins. In contrast, photorespiratory genes were mostly upregulated. Also, some genes of the Calvin cycle showed a mild transcript increase in *glk*1 *glk*2.

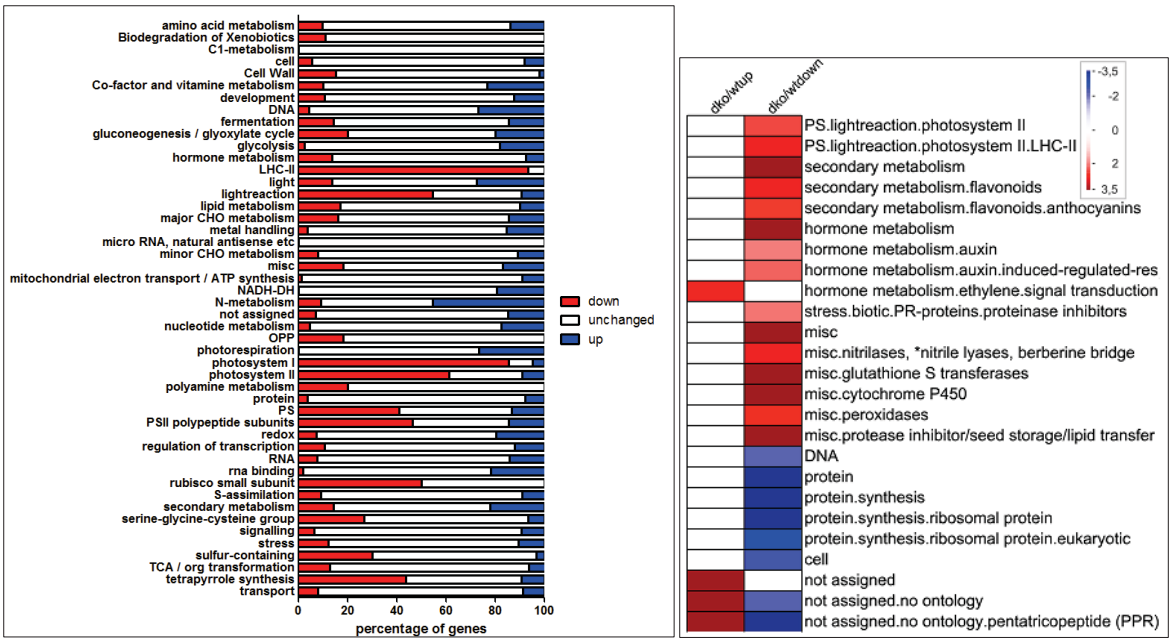


Figure 11: Bin classification of genes that are down-regulated (down) or up-regulated (up) in *glk1 glk2*. MapMan primary, secondary and tertiary bins such as a manually configured bin were used.

Figure 12: PageMan output. The figure is condensed to all MapMan bins where significant over-representation (red) or under-representation (blue) was found in either up- (left) or down-regulated (right) genes.

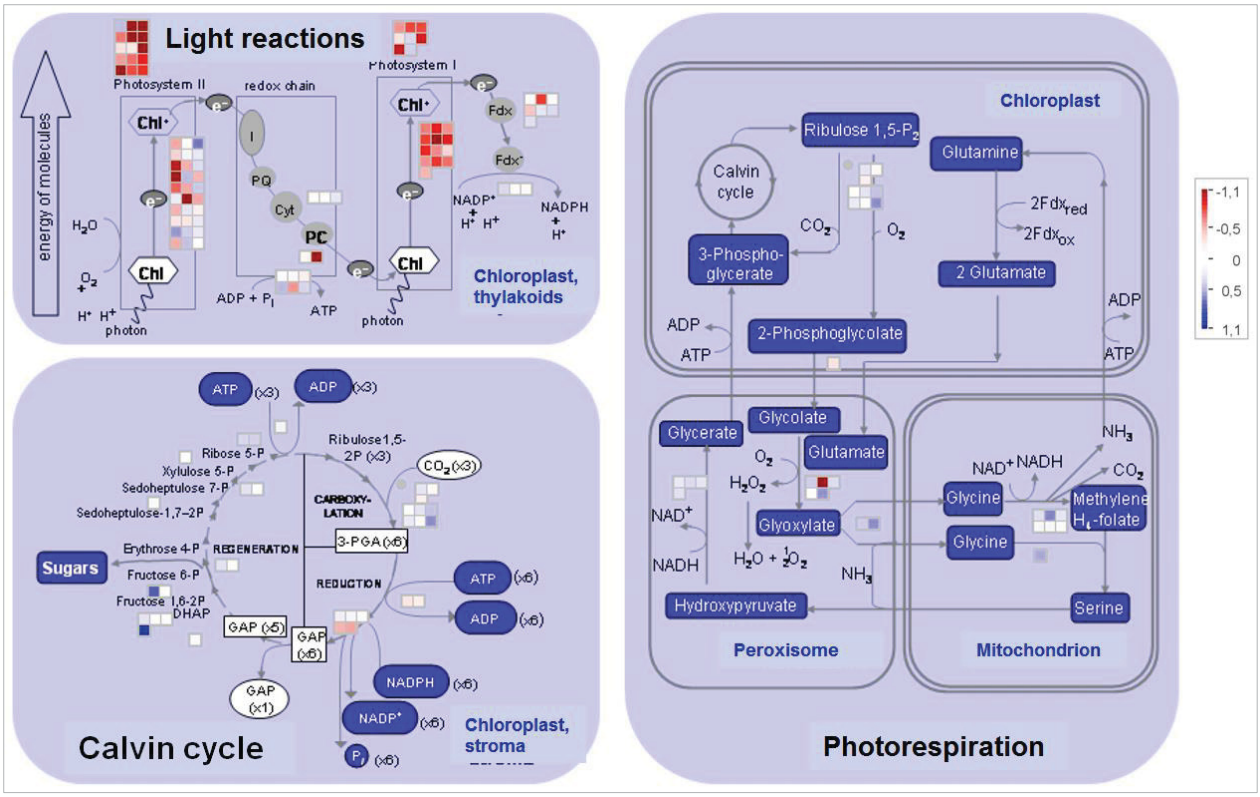


Figure 13: MapMan output. Differential expression of photosynthesis genes between wildtype and *glk*1 *glk*2 is visualized as colored squares. Red color indicates down-regulation in the mutant, blue color indicates up-regulation in the mutant, white squares represent unchanged genes.

Of the 110 genes that were differentially expressed between wildtype and *glk*2, 68 passed the filtering criteria. All but five of them were also de-regulated in *glk*1 *glk*2. Most genes in this group were assigned to the functional bins photosynthesis (11) and tetrapyrrole biosynthesis (7) (Figure 14, Supplemental dataset 3). Both PSI (4) and PSII (7) genes were among the down-regulated genes. The average decrease in expression was -26% for genes in this group.

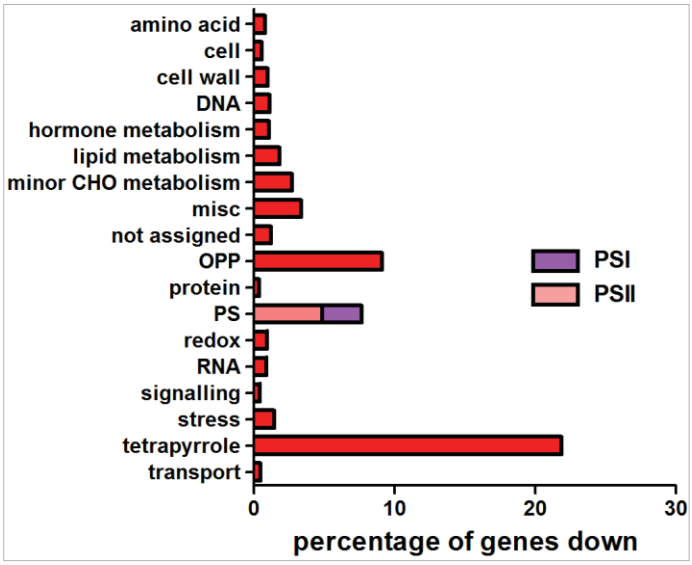


Figure 14: Genes that are down-regulated in glk2 assigned to primary MapMan bins. The PS bin is split into genes from PSI and from PSII. Four up-regulated genes were not included in the analysis, but can be seen in supplemental dataset 3. Bins with no matching genes in *glk*2 were excluded from the figure.

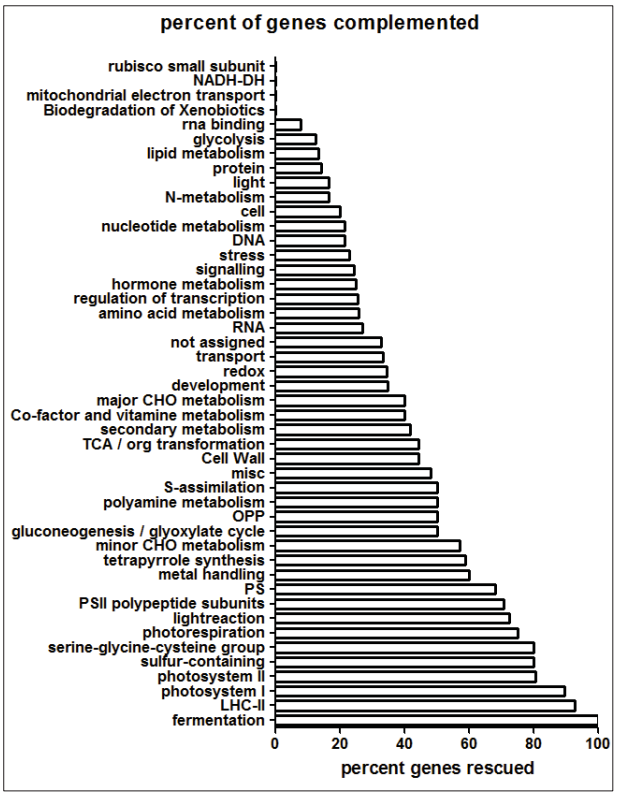
Photosynthetic genes are rescued with variation between functional groups

Next we quantified the relative degree of complementation of affected functional classes of genes in the complemented lines. We used two measures for complementation, the relative amount of genes that were complemented and the relative percentage of gene expression that was rescued for these genes. Groups of genes that are involved in photosynthesis uniformly showed high percentages of genes complemented with values ranging between 68% in the primary PS (photosynthesis) bin and 93% in the light harvesting complex II (LHCII, Figure 15). An intermediate rescue was monitored for genes in sulfur-assimilation, tetrapyrrole biosynthesis and secondary metabolism. Rarely any genes were complemented in glycolysis, light sensing and lipid metabolism. Mitochondrial NADH-dehydrogenase complex genes are up-regulated in *glk1 glk2* and remain uncomplemented completely. Our second measure, the rescued percentage of gene expression, was calculated as the

expression difference between co2.1 and *glk1 glk2*, normalized by the expression difference between *glk2* and *glk1 glk2*.

$$rescued\ percentage\ of\ expression = \frac{(expression\ in\ co2.1 - expression\ in\ glk1glk2)}{(expression\ in\ glk2 - expression\ in\ glk1glk2)} * 100$$

The average expression-deficit was rescued by 95.5%, with some variation between functional groups (Figure 16). Photosynthetic genes, in contrast to the high number of genes that are rescued, are among the least complemented genes in terms of expression values. Particularly low values are displayed by genes of the photosystem I (60,3%) and photosystem II core complex (65.0%). Full complementation of expression was achieved for genes of the light-harvesting complex II (95%), photorespiration (101%) and tetrapyrrole synthesis (114%).



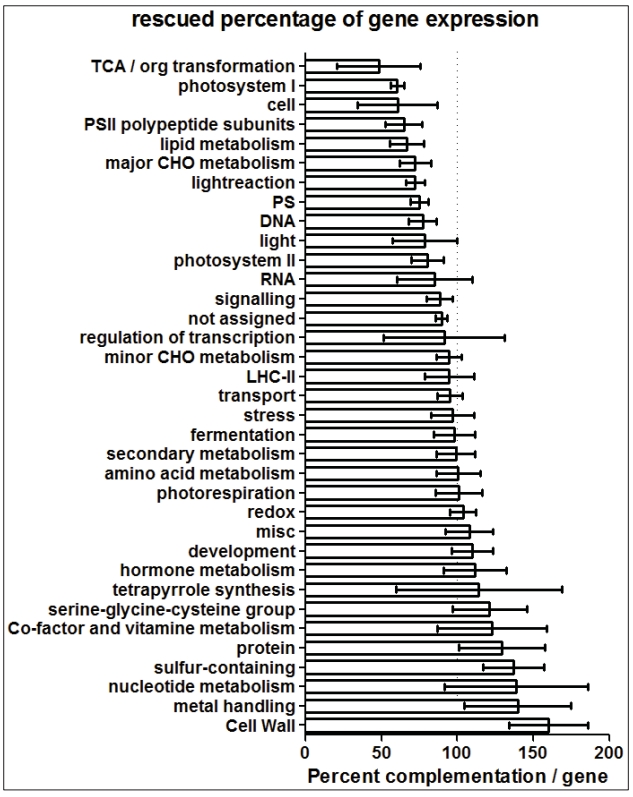


Figure 15: Percentage of genes that are complemented separated by bins. Bins were sorted from low to high percentages of complemented genes.

Figure 16: Percentage of gene expression that got rescued in complemented genes as mean +/- SE. Bins without complemented genes were excluded from the graph.

*Cg*Glk1 rescues light harvesting complex II genes significantly better than PSII core genes

We further investigated the heterogeneously complemented group of photosynthetic genes by comparing relative complementation of expression levels between signature genes (expression over 1000 RPKM) of the OEC and the light harvesting complex. All genes were significantly down-regulated

in *glk1 glk2*. Genes of the OEC were complemented by 37.4%, while LHC genes were complemented by 94.7% (Figure 17). We found these groups to be statistically different with a p-value of 0.034.

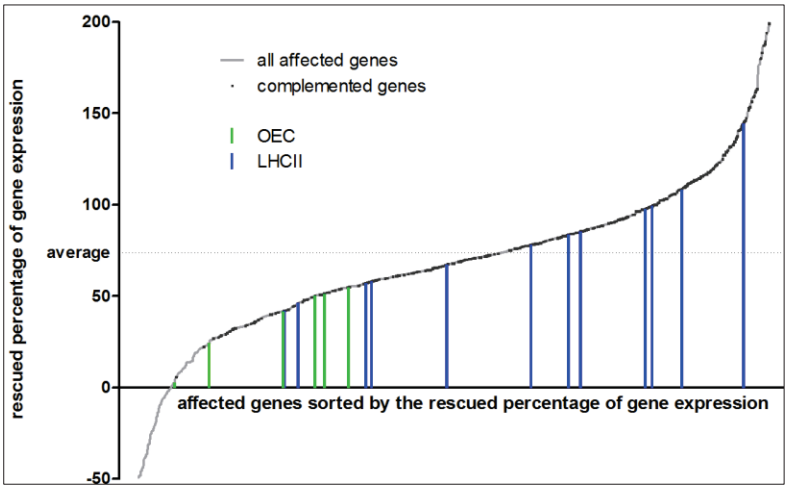


Figure 17: Rescued percentage of gene expression for all genes that were differentially expressed in *glk1 glk2*. Genes are sorted by ascending values. Genes with significant complementation are indicated as dark grey spots. Genes of the OEC and LHC of PSII are highlighted as bars.

Complementation of photosynthetic performance depends on the applied light intensity

To assess physiological defects in mutant plants, we tested functional parameters in all lines and compared them to the wildtype. As a proxy for plant fitness, we compared leaf photosynthesis rates at different light intensities. The light dependent plot of photosynthesis uniformly describes a saturating curve that reaches the maximum at 1000 to 1500 uE and starts decreasing at 2000 uE for all samples (Figure 18A). Photosynthetic rates were very similar when leafs were illuminated with 100 uE (variation < 6% from wildtype) except from glk1 glk2, which assimilated 15% less CO_2 than the wildtype (Figure 18B). When samples were illuminated at higher light intensities of up to 2000 uE, only *glk2* showed photosynthetic rates comparable to the wildtype (Figure 18A). The double mutant was impaired most with a decrease in net photosynthesis of 23% at 1500 uE. This difference was reduced to 11% or 14% in co1.1 and co2.1, respectively. In *glk1*, the impairment was similar to the complemented lines and reached 14%. In a statistical F-test, the curves of *glk1 glk2* and both complementing lines showed significance when compared to the *glk2*.

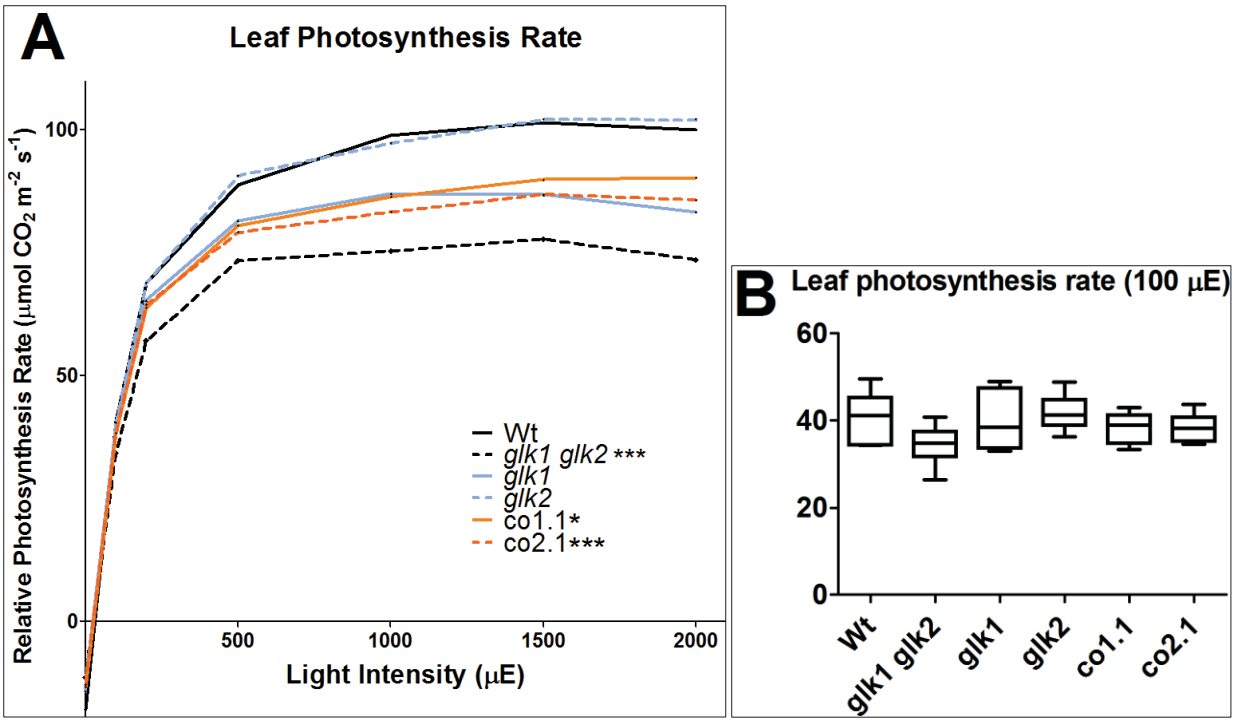


Figure 18: Relative photosynthesis rates in dependence of illumination were determined by gas exchange measurements. Five to six replicates of each genotype were measured. Asterisks indicate significance of a curve to the curve of *glk2* as tested by F-test, SD not shown (A). Visualization of photosynthesis rates at standard growth illumination of 100 µE (B).

To detect putative perturbations in photoprotection, we measured the maximum quantum yield of PSII and NPQ during illumination with high light. Photosystem II quantum yield was determined by the Fv/Fm ratio. All plants displayed Fv/Fm values between of 0.78 and 0.81 in a dark adapted state before the experiment (Figure 19). With increasing incubation time Fv/Fm values decreased, reaching 0.53 in wildtype leaves after 100 min. The quantum yield in the mutant decreased to 0.37 in the same time. While no complementation was seen in co1.1 (Fv/Fm of 0.34), co2.1 plants displayed a mildly rescued Fv/Fm value of 0.44 at the end of the experiment. The measurement of non-photochemical quenching (NPQ) was performed for ten minutes at illumination with 825 uE followed by ten minutes dark incubation. We monitored similar NPQ values for wildtype, *glk1* and *glk2*, peaking at 1.63, 1.68 and 1.72 at the end of illumination, respectively (Figure 20). NPQ was severely increased in *glk1 glk2*, peaking at 2.38 after ten minutes of high light treatment. The complemented lines, too, showed higher NPQ than the wildtype, where co1.1 peaked at 1.94 and co2.1 at 2.61. The degree of complementation we measured in chlorophyll fluorescence assays was less significant than complementation of visible features.

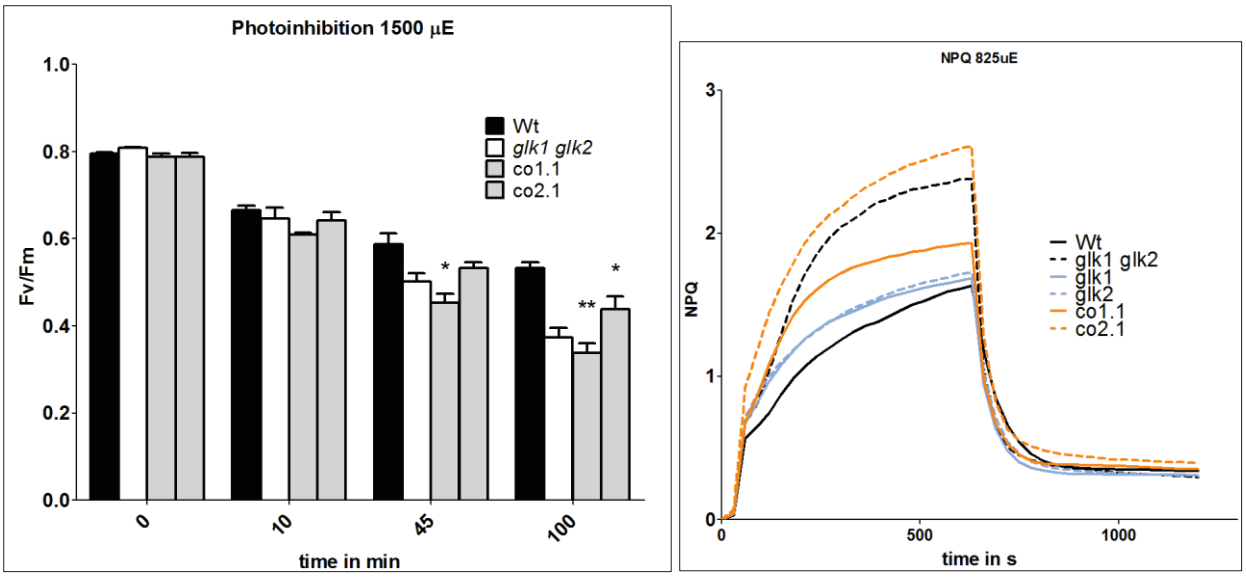


Figure 19: Fv/Fm values were monitored from leaf disks of dark adapted plants (0) and leaf disks after increasing incubation with high light (1500 uE). Fv/Fm was determined for three unique replicates per time point.

Figure 20: NPQ of wildtype, *glk* mutants and complementing lines. NPQ was measured on intact leaves of dark adapted plants at 825 µE for 10 minutes. Recovery was measured for 10 minutes in darkness.

Metabolite levels are rescued by complementation with *Cg*Glk1

To analyze, how well metabolic processes are rescued by complementation, we measured the relative abundance of 41 metabolites by GC/MS analysis of leaf extract. Ten out of 41 metabolites were significantly changed in *glk1 glk2*, all of them showed lowered abundance (Figure 21). Among these metabolites, we found three intermediates of central carbon metabolism (citrate, glucose, maltose), three amino acids (glycine, serine, lysine), as well as glycerate, glycerol, lactate and glycolate. The latter three metabolites are intermediates of lipid biosynthesis, secondary metabolism and photorespiration, respectively. The decrease in abundance was most significant in glycine, serine, maltose and glycerol, where over 50% of the respective metabolite was lost in the mutant. Abundance of all metabolites was rescued in both single mutants and complementing lines. Three of the ten metabolites (serine, glycine, glycerate) are intermediates of reactions that are represented by the bin “amino acid metabolism.synthesis.serine-glycine-cysteine group”, which we found down-regulated in glk1 glk2 and rescued in complemented lines in our RNA-sequencing data. Cysteine, the third main metabolite of this biosynthesis cluster, was not significantly decreased in *glk1 glk2* but showed a similar trend.

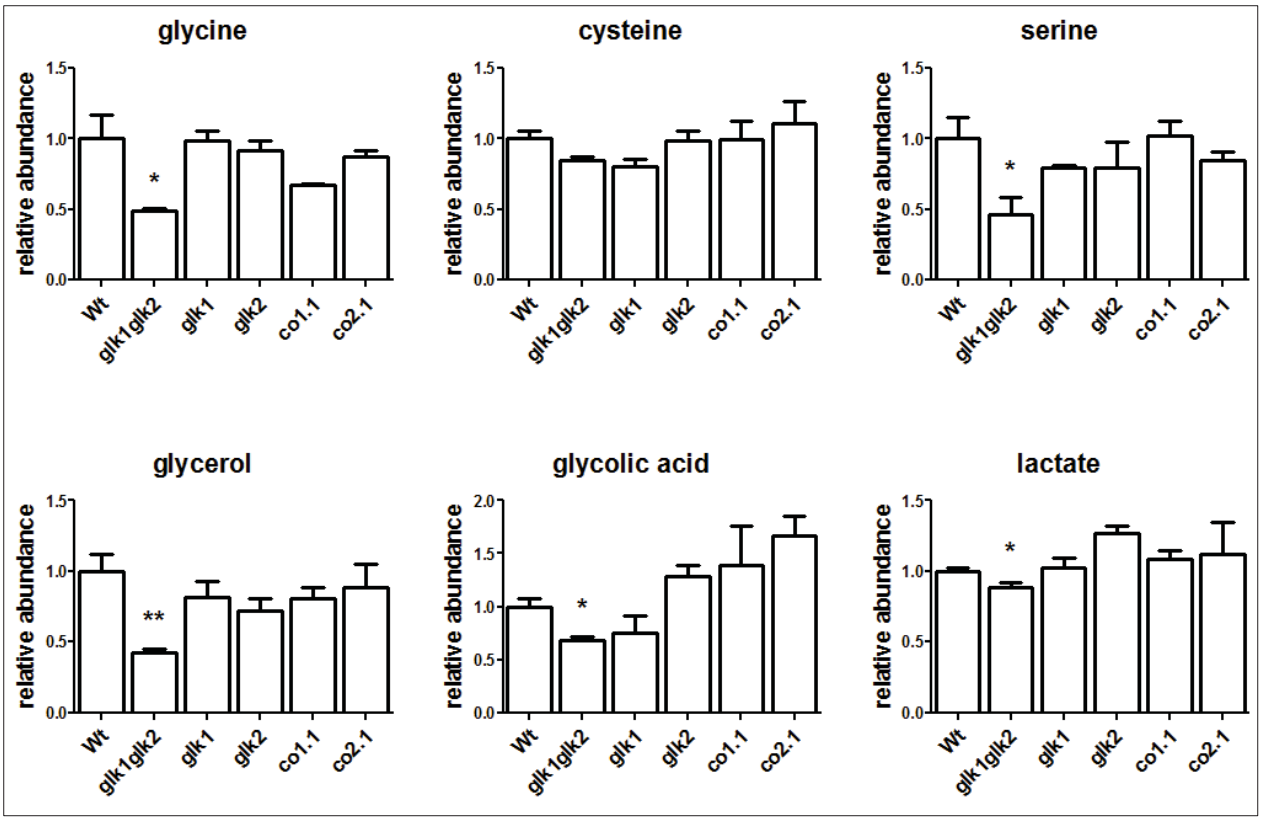


Figure 21: Metabolite abundance in leaves normalized to wildtype levels. Five out of ten significantly changed metabolites and cysteine are showed. Significance is indicated with asterisks, where *=p<0.05 and **=p<0.01.

Discussion

In this study, we demonstrate that the Glk genes from *A. thaliana* do not act fully redundant in leaves. Loss of a single Glk gene, however, does not cause a physiologically detectable phenotype. With deep RNA sequencing, we further show that mutation of both *glk*1 and *glk*2 leads to de-regulation of 8.8% of all transcripts. A phenotypic rescue by *Cg*Glk1 in the mutant indicates at least partial conservation of Glk1 function between the C_3 plant *A. thaliana* and its close relative C_4 species *C. gynandra*. Interestingly, we found an unequal rescue of genes of PSII, where OEC transcripts were rescued to lower extent than LHC genes. Future analyses with overexpressor lines will help to determine, whether this effect is a result of molecular evolution of *Cg*Glk1.

Phylogenetic analysis and multiple alignment of Glk protein sequences

In our study, we complemented *glk1 glk2* mutants of *A. thaliana* with the Glk1 gene from *C. gynandra*. Since the *C. gynandra* genome is not sequenced, we derived sequence information from RNA sequencing data. To test whether Glk genes from both species are orthologs, we conducted phylogenetic analysis. Glk1 proteins from *Brassicales* were distinct from Glk2 proteins of the same species (Figure 1), which shows that these duplicated Glk genes origin from a single genome

duplication event and therefore represent orthologs. A shared genome duplication event between *Cleomaceae* and *Brassicaceae* within the *Brassicales* was proposed earlier (Barker et al., 2009). The Common origin of *Cg*Glk1 and *At*Glk1 is a strong indicator of functional similarity of these genes. We compared the amino acid sequences Of *C. gynandra*, *T. hassleriana* and *A. thaliana* in a multiple sequence alignment to analyze molecular similarity (Figure 2). The DNA binding helix-loop-helix domain was highly conserved, which is another indicator for functional conservation of Glk1 between the closely related C_3 and C_4 species. Surprisingly, the GCT-box, which was shown to be essential for protein-protein interactions in maize Golden2 (Rossini et al., 2001; Fitter et al., 2002), showed eight amino acid exchanges between *A. thaliana* and *C. gynandra* and large parts of it were lost in *T. hassleriana*. The divergence of this conserved domain in *Cleome* species is in contrast to our initial hypothesis of conservation of Glk1 function throughout the *Brassicales* and independent of the photosynthesis type. Previous studies showed that single amino acid substitutions can cause a partial or complete change of targets of a transcription factor (Hsia and McGinnis, 2003). It was further hypothesized that a change in protein-protein interactions of a transcription factor suits well as an initial step of evolution of its target specificity (Tuch et al., 2008). As Glk1 is the only expressed Glk family member in *C. gynandra*, changes in its affinity to other proteins might be a result of adaptation to this secondary loss of Glk2 function.

_Cg_Glk1 localizes to the nucleus of leaf tissue cells

One part of this study was the analysis of *Cg*Glk1 expression in the *glk*1 *glk*2 background in *A. thaliana*. To make predictions on the tissues that express *Cg*Glk1, we performed localization studies with a GUS protein fused to the *Cg*Glk1 promoter. The construct showed expression in all cell types of leaf tissue (Figure 5). This result points out that the heterologous promoter activates gene expression in bundle sheath and mesophyll cells of *A. thaliana* and is therefore a suitable promoter for complementation studies in this C_3 plant. A GFP tagged version of the *Cg*Glk1 gene stained a single, prominent organelle that was identified as the nucleus in *N. benthamiana* protoplasts, which indicates nuclear targeting of the gene (Figure 6). Since Glk1 is a DNA-binding transcription factor, it is thus likely to be functional in *A. thaliana*, based on its localization.

Molecular analysis of the _glk1 glk2_ mutant: RNA-seq of nuclear transcripts

To elucidate the impact of *glk*1 *glk*2 mutation on the transcriptome and the ability of *Cg*Glk1 to complement this phenotype, we conducted RNA-sequencing of mRNA of nuclear encoded genes on one lane of an Illumina HighSeq 2500 flow-cell. The quality of the run was satisfying, since 94% of 196 million reads could be mapped to the *A. thaliana* transcriptome (Table 1). Similarity analysis of the samples showed that the single mutants *glk1* and *glk2* are very similar to the wildtype as only 0.1%

and 0.5% of their transcriptome was significantly changed, respectively (Table 2). In *glk1 glk2*, 8.8% of all genes were differentially expressed. It therefore showed the highest observed difference to the wildtype, which was expected due to the severe phenotype of the double mutant. In the complementing lines co1.1 and co2.1, 6.3% and 3.2% of all genes are differentially expressed, respectively. This aligns well with the expression strength of *Cg*Glk1 of 13.5 RPKM in co1.1 and 38.0 RPKM in co2.1 (Figure 9) and demonstrates that complementation is a quantitative effect. The pairwise comparison of the complementing lines with *glk1 glk2* also indicates better complementation in co2.1 with 0.7% differentially expressed genes in co1.1 and 2.7% in co2.1 (Table 2). We performed a principal component analysis (PCA) of gene expression values in all replicates and found two components that together explained 31% of all variation in the dataset (Figure 7). The first component explained 22% of all variation and separated the replicates of all genotypes with the extremes being wildtype on one side and *glk1 glk2* on the other side. Interestingly, we observed slight overlaps between replicates of wildtype and *glk1* and between replicates of *glk1* and *glk2*. This is another indicator for the high similarity of the transcriptomes of single mutants and wildtype. Again, we found co2.1 to be closer to wildtype than co1.1, which was closest to *glk1 glk2*. The second component explained 9% of all variation but we observed no systematic separation of replicates of different genotypes. The results of PCA were complemented by a hierarchical clustering (HCL) analysis (Figure 8). The separation of phenotypically impaired plants (*glk1 glk2*, co1.1, co2.1) from phenotypically mostly healthy plants (Wt, *glk1*, *glk2*) was distinct, while replicates could not be separated between wildtype and glk1 and between the complementing lines. For wildtype and *glk1*, this inadequacy results from too few transcriptional differences under the tested conditions. The high similarity of the complementing lines despite significant changes in other experiments can be explained by the tendency of Pearson correlation to correlate relative rather than absolute values. Since the focus of this analysis was on predominantly highly expressed photosynthesis genes, we decided to introduce a significance cut-off of an average expression of >20 RPKM in at least one genotype for further analyses. We found validation for this cut-off value by achieving a 2.38-fold increase in differentially expressed genes and at the same time ruling out 97% of noise of organelle-encoded genes.

We analyzed differentially expressed genes of all genotypes with visual impairment as compared to the wildtype and summarized the results in Venn diagrams, separated by up- and down-regulated genes (Figure 10). The largest portions of genes are either only de-regulated in *glk1 glk2* or shared between *glk1 glk2* and at least one other line. For the complementing lines, this visualizes a partial rescue of the mutant phenotype. The portion of common de-regulated genes between *glk1 glk2* and co1.1 is larger than the shared portion between *glk1 glk2* and co2.1. This indicates that genes that are rescued in co1.1 often are also rescued in co2.1, while the opposite is not true. This is in line with the higher expression of *Cg*Glk1 in co2.1, which allows for better complementation of de-regulated genes.

A very small number of genes, namely 15, are de-regulated in both complementing lines and not in *glk1 glk2*. This group does not contain any known C_4 genes and the change of transcript levels is only 37% (data not shown). It is thus very likely that these genes and the 151 genes that are deregulated only in one complementing line are due to individual secondary effects upon different shades of functional complementation of the double mutant. When comparing de-regulated genes from *glk*2 with *glk*1 *glk*2, the large overlapping fraction of genes indicates that *glk*2 indeed displays a mild version of the double mutant's phenotype on the molecular level (Figure 10). This observation was surprising, as previous studies described *At*Glk1 and *At*Glk2 as functionally redundant (Yasumura et al., 2005; Waters et al., 2008; Waters et al., 2009). A major fraction, namely 18 of the 62 shared, down-regulated genes between *glk*2 and the double mutant, were assigned to photosynthesis and tetrapyrrole biosynthesis (Figure 14). This fits well with the description of these genes as targets of Glk regulation in Arabidopsis (Waters 2009). Surprisingly, the group of down-regulated photosynthesis genes contained not only genes that were described as Glk targets by Waters and colleagues (At3G54890, At3G27690, At2G05070, At2G34430, At1G29930, At2G05100), but also additional proteins of LHC and core complexes (At1G19150, At3G16140, At2G20260, At1G29910, At4G15510). We also found five further tetrapyrrole biosynthesis genes decreased in *glk*2 (At1G08520, At5G08280, At1G03630, At4G01690, At2G40490) (Supplemental dataset 3). The increased number of genes that were detected in this experiment might be either due to higher sensitivity of RNA sequencing over microarray analysis (Zhao et al., 2014) or result from indirect, secondary effects on photosynthesis genes upon transcript decrease of Glk targets. Single knockouts of tetrapyrrole biosynthesis genes were demonstrated to negatively affect expression of their precursors in the pathway by an unknown mechanism (Tanaka and Tanaka, 2007). In *glk*2, however, tetrapyrrole biosynthesis is not completely shut down but rather slightly reduced as indicated by the green color of the plant. Also, the affected genes code for enzymes that do not obtain consecutive functions in tetrapyrrole biosynthesis. Our observations thus make a direct targeting of these genes by Glk and, as a consequence, a transcript decrease in *glk* mutants more likely.

To analyze differences in the *glk1 glk2* transcriptome in a functional context, we classified all genes by their MapMan annotation and compared them for enrichment in de-regulated genes. All primary MapMan bins, some manually selected secondary and tertiary bins were included to be able to adequately illustrate differential expression in photosynthesis genes. It is thus possible that genes are counted multiple times, in their primary, secondary and tertiary MapMan bin. The highest percentage of differentially expressed genes between *glk1 glk2* and wildtype was found in functional classes related to photosynthesis, where almost exclusively genes were down-regulated in *glk1 glk2* (Figure 11). This observation is in line with a previous study, where microarray data of *A. thaliana glk1 glk2*

mutants that were complemented with inducible *At*Glk constructs were analyzed (Waters et al., 2009). Among the 30 most responsive genes, Waters and colleagues found 13 to be involved in photosynthesis or chlorophyll biosynthesis. While only 41% of all genes in the primary bin PS (photosynthesis) are down-regulated, this number increases to 71% in photosystem II, 91% in photosystem I and 93% in LHCII. This complements the finding that genes of the photosynthetic apparatus are the central targets of *At*Glk genes (Waters et al., 2009). We can, however, not conclude, whether the decrease of this broad range of photosynthesis proteins is a direct or indirect effect of a lack of Glk. Accordingly, 53% of all genes in tetrapyrrole and therefore chlorophyll biosynthesis were disturbed in expression. Here, we also found all genes that were decreased in *glk*2, which makes tetrapyrrole synthesis a common target of glk mutants. Interestingly, many bins that are not directly related to photosynthesis showed numerous de-regulated genes with the highest percentage in nitrogen metabolism, where over 50% of all genes ore de-regulated. Also carbohydrate, lipid and aminoacid metabolism are affected with a particularly strong effect on the synthesis of serine, glycine and cysteine. This is well aligned with a decreased abundance of glycine and serine in *glk*1 *glk*2 (Figure 21). These massive changes in primary metabolism show that a loss of both *glk*1 and *glk*2 mediates a widespread phenotype throughout plant metabolism. Regardless of whether the broad de-regulation is a direct or indirect effect of Glk mutation, this result reveals the global impact of Glk expression. The functional classes with the highest percentages of up-regulated genes in *glk1 glk2* are photorespiration, N-metabolism and light signaling. An increase in photorespiration, the salvaging pathway of RuBisCO oxygenation, is probably a result of heavy disturbance of expression of the photosynthetic apparatus and RuBisCO in *glk1 glk2*. The increase in N-metabolism is most likely a combination of a secondary effect that enhances refixation of ammonia that was lost during photorespiration (Wallsgrove et al., 1983) and a feedback response to the disturbed photosynthetic apparatus. The latter is probably also true for an up-regulation of light signaling.

To complement the enrichment analysis, we conducted over-representation analysis of de-regulated genes with PageMan. The over-representation of photosynthesis genes among the down-regulated genes is rated significant in this analysis. Interestingly, we found a group of bins over-represented in down-regulated genes that share a function in auxin metabolism (Figure 12). A decrease in genes controlling this central growth hormone is a molecular indicator for the retarded growth of the *glk1 glk2* mutant. As a third group of over-represented bins among down-regulated genes, we found anthocyanin metabolism. Since we monitored an increased oxidative stress of *glk1 glk2*, accumulation of anthocyanins are likely part of a photo-oxidative-stress response. Anthocyanins were shown to be involved in ROS scavenging (Chalker-Scott, 1999; Gould et al., 2002). The group of under-represented bins among the down-regulated genes mainly consisted of the terms DNA and protein. These findings, again, are in agreement with the previous analysis. However, despite the high

percentage of genes that are up-regulated in the bin “N-metabolism”, PageMan analysis did not find this to be significant. Over-representation analysis of up-regulated genes found only a single group of proteins enriched that is not assigned any function. This stands in contrast to the equal number of genes that are up- or down-regulated in *glk1 glk2*. Glk genes are transcriptional activators and a lack of Glk proteins thus causes down-regulation of their targets (Waters et al., 2009). We suppose that only primary effects cause enrichment of de-regulated genes in functional groups, as the high number of genes from various bins that are affected by secondary effects upon *glk* mutation dilutes visible effects (Waters et al., 2009).

To study expression of the main targets of Glk regulation, namely photosynthesis genes, in detail, we visualized expression changes gene-wise with the MapMan tool (Figure 13). Here, the decrease of transcripts in light reactions can also be seen for plastocyanin and some ferredoxins, demonstrating that electron-shuttling proteins are, too, involved in the response to a lack of Glk. This tool also addresses genes to functions in the calvin cycle, which was not achieved by the previous categorization tools. Interestingly, calvin cycle genes are not down-regulated in *glk*1 *glk*2. Some enzymes even display slight up-regulation. As the calvin cycle fixes carbon and is therefore a key pathway of biomass generation, this response stands in contrast to the uniformal decrease in photosynthetic transcripts and the associated dwarfish phenotype. Since calvin cycle enzymes are strongly regulated on the posttranslational level (Michelet et al.), it is likely that adaptation to a decreased energy production happens on mature proteins rather than mRNA. However, this shows that *glk*1 *glk*2, despite being restricted in energy production, invests energy and resources to maintain this consumptive pathway at wildtype level, thereby highlighting its essential role in plant metabolism. We conclude that there is no signaling cascade that regulates calvin cycle gene expression in response to performance of light reaction, as light fluctuation in a natural environment occurs in a manner of seconds and makes calvin cycle genes not suitable for transcriptional regulation. The glk1 glk2 mutant provides an ideal environment to observe this lack of response of the calvin cycle, since reduction of light reaction genes is severe, but still enables functional photosynthesis.

Phenotypical complementation of the *A. thaliana glk1 glk2* mutant by *Cg*Glk1

To test on functional conservation of Glk1 genes, we transformed the *A. thaliana glk1 glk2* double mutants with *Cg*Glk1 (Figure 3). The resulting plants showed full complementation of the visual phenotype when Glk1 was driven by a strong 35S promoter, thereby indicating conservation of Glk function in the order of Brassicales. Functional conservation was demonstrated for Glk genes between *A. thaliana* and the moss *Physcomitrella patens* earlier, indicating that conservation of Glk function is widespread in land plants (Yasumura et al., 2005). The native Glk1 promoter from *C. gynandra*, however, was not able to drive *Cg*Glk1 expression towards full complementation of the visual

phenotype. This is probably due to weak expression of the construct, which may be a result of the evolution of C_4 photosynthesis in *C. gynandra* and changed properties of the *Cg*Glk1 promoter. C_4 directed evolution of promoters was shown for the dicotyledonous C_4 plant *Flaveria bidentis* (Engelmann et al., 2008). Overall expression of *Cg*Glk1 in co1.1 and co2.1 was significantly lower than AtGlk1 expression in wildtype plants (9 and 25% of *At*Glk1 wildtype expression, respectively). Expression values of co2.1 were thus taken as a reference for the degree of complementation.

The chlorophyll content of the 35S promoter-driven, complemented line co2.1 was, in contrast to the plants' appearance, not restored completely but rather to 86% of the wildtype level (Figure 4). Surprisingly, the *glk2* single mutant shows an equal loss of chlorophyll as compared to the wildtype. The impaired expression of some photosynthetic transcripts that we already monitored earlier is most likely causal for this mild chlorophyll-content phenotype. In previous studies, a pale-green phenotype of *glk2* that was only visible in siliques led to the hypothesis of redundancy between the Glk genes in other tissues of *A. thaliana* (Waters et al., 2008). Our results show that also green leaf tissue is affected by the loss of a single Glk, which is well in line with our hypothesis of an incomplete functional redundancy of Glk genes in *A. thaliana*. It was shown that *glk1 glk2* mutants, which were complemented with the 35S promoter-driven Glk1 gene from *A. thaliana*, contained more chlorophyll than the wildtype (Waters et al., 2009). This complementation due to overexpression may be explained by a gene dosage effect that helps to overcome the decreased chlorophyll-content (Fitter et al., 2002). The insufficient complementation of chlorophyll levels in the complementing lines indicate that gene expression might not be fully restored.

Complementation of the molecular phenotype by *Cg*Glk1

Here, we analyze the results of RNA sequencing of the complemented lines to assess the potential of *Cg*Glk1 to rescue the de-regulated transcriptome from *glk*1 *glk*2. As a measure for complementation of functional classes of genes, we calculated the percentage of genes in each bin that is complemented in at least one of the complementing lines, co1.1 and co2.1 (Figure 15). Since full complementation of Glk1 transcript level was not achieved in the complementing lines and the expressed protein is from a heterologous organism, we focused the complementation analysis on these qualitative observations. The most thorough complementation was seen in photosynthesis bins, which demonstrates general functional conservation of Glk1 function between *C. gynandra* and *A. thaliana* on transcriptional level (Waters et al., 2009). Genes that are involved in chlorophyll biosynthesis are included in the class tetrapyrrole biosynthesis, where also many genes are complemented. This aligns well with the complementation of chlorophyll levels in the complementing lines and again shows functional conservation between *Cg*Glk1 and *At*Glk1. Classes that were assumed to be de-regulated by secondary effects, like photorespiration, show a high relative number of complemented genes (75%)

as well. This indicates that photosynthesis in the complemented lines, at illumination with moderate light, is largely rescued from a qualitative point of view. Among the well complemented groups we further found secondary metabolism of sulfur-containing compounds and the synthesis pathway of serine, glycine and cysteine, which illustrates that some non-photosynthetic pathways were also rescued by *Cg*Glk1 expression. On the other hand, two thirds of all de-regulated genes were not complemented and according to categorization, they were found in functional bins that do not contain known Glk targets and are therefore most likely de-regulated due to secondary effects upon loss of Glk. We found a low percentage of complemented genes in lipid and protein metabolism, also redox and stress genes remained un-complemented. The widespread failure of *Cg*Glk1 to complement secondary effects, led us to the hypothesis that the comparatively low expression of *Cg*Glk1 is not sufficient to restore expression of photosynthetic genes to wildtype level. To test this hypothesis, we quantified the rescued percentage of expression of each gene that was significantly complemented. For this calculation, we used the expression values of co2.1, since it was the stronger complementing line (Figure 16). Here, we found that most complemented genes of photosynthesis-related groups were far from complete rescue of expression (100%). Bins containing genes that were affected by secondary effects, however, showed high relative complementation. Sometimes complementation exceeded 100%. This was due to very few outliers with extremely high values of relative complementation, which is reflected by high standard errors in these bins. The average degree of rescue in a bin was 95.5%. To understand the insufficient complementation in the complementing lines, we further focused only on photosynthesis genes in this analysis. The rescued percentage of expression was only 60% and 65% for PSI and PSII core genes, respectively, which proves our hypothesis that the complementation of some photosynthesis genes is incomplete and thus, secondary effects can still be observed in complementing lines. In contrast to core genes of the photosystems, gene expression of LHCII genes was rescued by 92.0%. The gap of 27% between two groups of the same protein complex, namely PSII, brought us to the question, whether different groups of genes might display different affinities to *Cg*Glk1. To study the unequal complementation of PSII genes in more detail, we manually screened all genes in the respective MapMan bins that were de-regulated in *glk1 glk2* (Figure 17). We found 13 signature genes (expression over 1000 RPKM) in the LHCII bin and 11 signature genes in the PSII core bin. Of the PSII core signature genes, 6 were the main members of the OEC of PSII: PsbO1+2, PsbP, PsbQ1+2 and PsbR. These six genes were strongly impaired in *glk1 glk2* and their relative rescue of expression was only 37.4% in co2.1, which was less than any recorded rescue in a MapMan bin. In contrast, the signature genes of LHCII were rescued by 94.7%, which is very close to the previously calculated value for LHCII rescue. This is because only one LHCII gene was excluded from analysis for not being a signature gene and the datasets are thus similar. In maize, a similar reduction of expression of OEC genes was detected in an

earlier study (Majeran et al., 2008). We tested the values for rescue of gene expression in LHCII and OECII for independence with a T-test and found statistical significance with p=0.034. This result demonstrates that *Cg*Glk1 complements genes of the OEC in a different manner than LHCII genes and is well in line with our earlier findings. Full complementation of the decreased chlorophyll content is probably due to the complete restoration of LHCII expression, since LHCs contain a large percentage of cellular chlorophyll (Thornber, 1975). We do not suggest a loss of function of excess LHCII proteins, as a similar imbalance between low PSII core and high LHCII genes has been shown in BS cells of maize wildtype, too (Majeran et al., 2008). It was also hypothesized that LHCII genes are recruited to PSI core complexes (Majeran 2008). Since a different composition of photosystems is supposed to have an effect on photosynthetic parameters, we measured photosynthesis and photoinhibition in saturating light regimes to complement our findings.

Functional complementation of *glk1 glk2* by *Cg*Glk1 is incomplete

The photosynthesis rates of mature leaves were analyzed to quantify functional defects in Glk mutants and the degree, to which they are complemented in co1.1 and co2.1. We measured similar photosynthesis rates of all lines at the standard growth light intensity of 100 µE, just *glk1 glk2* photosynthesis was 15% lower than the wildtype (Figure 18B). While the phenotypically severely impaired *glk1 glk2* mutant has a mild photosynthesis phenotype, gene expression in the single mutants and complemented lines is sufficient for photosynthesis at moderate light. These results demonstrate the sufficiency of a single Glk gene for *A. thaliana* to grow and they align well with a previous study, where functional redundancy was shown for Glk genes in *A. thaliana* (Fitter et al., 2002). At high light intensities, however, both complemented lines show a decrease in photosynthesis versus the wildtype (Figure 18A). The decrease is about half as strong as in *glk1 glk2*. This indicates that the composition of the photosynthetic apparatus that is mediated by *Cg*Glk1 causes insufficiency to process large quantities of photons. In contrast, the *A. thaliana glk2* mutant performs as good as the wildtype at high light intensities. This shows that the photosynthetic phenotype that is observed in the complementing lines does not only depend on a lack of Glk2 or on the relative degree of complementation as approximated by chlorophyll levels, but rather results from the incomplete rescue by mild expression of *Cg*Glk1. Interestingly, the glk1 mutant displayed retardation in photosynthetic performance that is close to those of the complementing lines. The most likely reason is that *A. thaliana* Glk1 is involved in the light-stress response although evidence for this hypothesis is lacking, since *glk* mutants were never characterized in high light. When all graphs were statistically tested versus *glk2*, differences were significant for *glk1 glk2* and both complementing lines according to an F-test. The *glk*1 curve, despite being very different from *glk2*, showed a p value of 0.08, which is most likely due to high standard deviation in high-light stressed plants and even N ≥ 5 did not resolve this

issue. We suggest that the reason for a photosynthetic phenotype in the complementing lines only upon high light illumination is that saturation of linear electron transport occurs at lower photosynthesis rates. This could putatively result from a decrease in the water splitting reaction of PSII (Allahverdiyeva et al., 2009).

High light intensities cause oxidative stress at the site of the PSII core to which plants react by deactivating PSII via degradation of the central D1 protein (Barber and Andersson, 1992). This process is called photoinhibition. We measured chlorophyll fluorescence to detect whether photoinhibition is disturbed in mutants and complementing lines. Chlorophyll fluorescence is directly related to the maximum quantum yield and therefore the activation state of PSII (Genty et al., 1989). The very similar Fv/Fm values in dark adapted plants of wildtype, *glk1 glk2* and the complementing lines show that no impairment of photosystem II occurs due to modified expression of Glk targets alone (Figure 19). We therefore conclude that PSII is functional per se in all examined plant lines and that a small quantity of photons can be processed by their residual core complexes. Upon illumination with high light (1500 µE) for 100 minutes, though, both complemented lines and *glk*1 *glk*2 display significantly lower Fv/Fm values than the wildtype. A decreased Fv/Fm value indicates increased photoinhibition of the plant, as it is also displayed by mutants of PSII core genes in *A. thaliana* (Murakami et al., 2005; Ishihara et al., 2007; García-Cerdán et al., 2009). This finding thus complements our previous results in showing that *Cg*Glk1 does not fully complement genes of PSII.

As a second complementary approach to determination of the maximum quantum yield of PSII, we measured the relative amount of non-photochemical quenching (NPQ) at high light illumination with 825 µE (Figure 20). While the single mutants glk1 and glk2 showed an NPQ similar to the wildtype, *glk1 glk2* and the complemented lines showed an increased NPQ. An increase in NPQ is an indicator for decreased photochemical quenching, which integrates well with previous results, where the double mutant and the complementing lines displayed a decrease in photosynthesis rate and an increased photoinhibition. We suggest that the reduced potential to drain excitation energy via PSII as a consequence of the reduced amount of core genes causes the increased NPQ. In co2.1, NPQ is even higher than in the *glk*1 *glk*2 double mutant. We conclude that the high abundance of LHCII genes paired with reduced levels of PSII core genes further increases NPQ, since more electrons get excited by photons, while photochemical quenching is still impaired and cannot quench the energy. Increased photoinhibition as seen in the Fv/Fm measurement is a second consequence of this imbalance. This result, too, fits well to our observation of an unequal rescue of PSII genes.

Conclusion

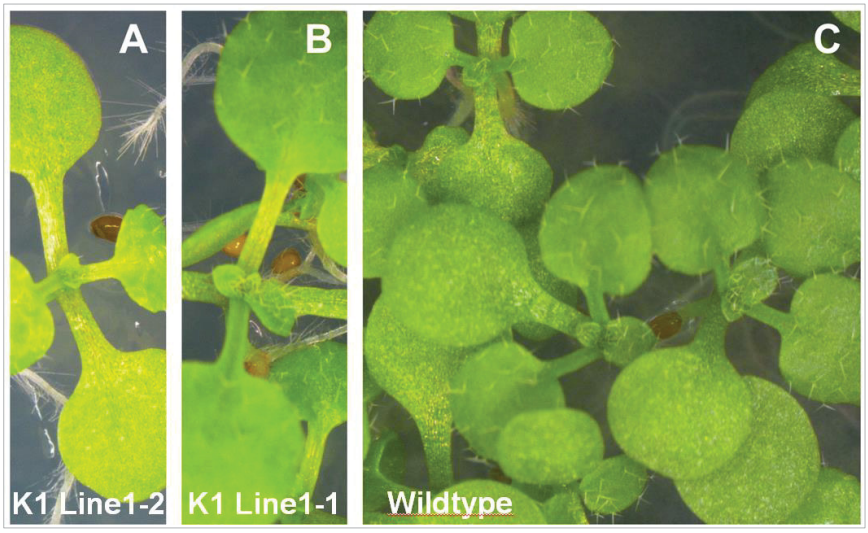
From the data presented in this work, we conclude that the pair of Glk transcription factors does not act completely redundantly in *A. thaliana*. The knockout mutant of *glk*2 displays a mild molecular and

physiological phenotype. By comparative RNA sequencing of wildtype and *glk* mutants, we found further photosynthesis genes to be regulated by Glk factors and revealed the global impact of the *glk*1 *glk*2 double knockout on the transcriptome. We showed that despite severe de-regulation of light reactions, the calvin cycle is unaffected on transcript level and conclude that a long term response to this artificial light stress is not activated or missing.

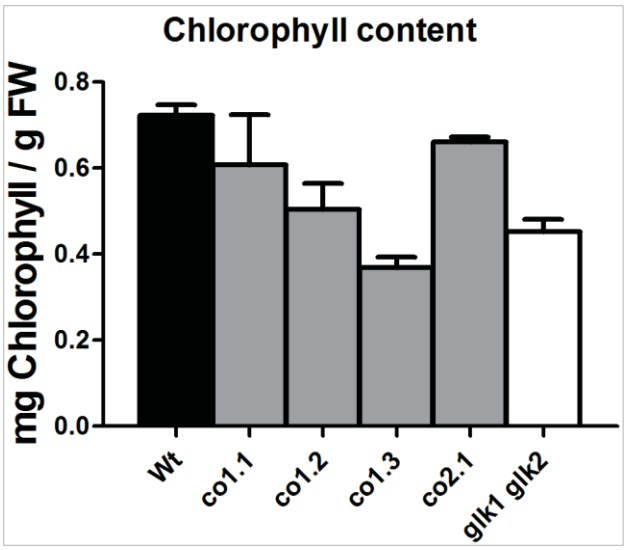
We further show that the function of Glk1 is conserved between *A. thaliana* and *C. gynandra*. Mild expression of *Cg*Glk1 complements the growth phenotype of the *A. thaliana glk*1 *glk*2 mutant, but does not suffice to reset all transcripts to wildtype level. Detailed analysis of photosynthesis genes, which are known Glk targets, showed that there are significant differences in the rescued amount of transcript between genes of the OEC and the LHC of PSII. This differential complementation might be a result of Glk1 evolution in *C. gynandra*. Finally, we complemented our findings with physiological studies that reflected changes in the transcriptome adequately.

To assess, whether *Cg*Glk1 lost the ability to activate transcription of certain photosynthetic genes, we will conduct a new screen for plants overexpressing *Cg*Glk1 and detect photosynthetic transcripts of OEC and LHC genes in these lines. We would thus rule out insufficient complementation as a result of a low gene dosis.

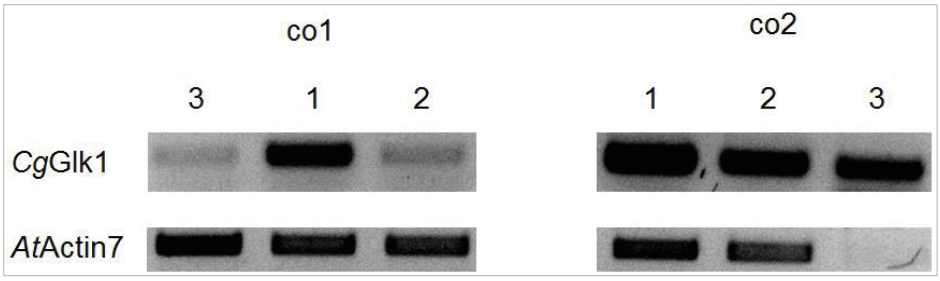
Supplemental Information



Supplemental Figure 1: Differential complementation of the pale green phenotype of *glk*1 *glk*2 by *Cg*Glk1 expression driven by the native *Cg*Glk1 promoter as compared to the wildtype. Seven day old seedlings were analyzed.



Supplemental Figure 2: Chlorophyll content of complemented plants. Three lines carrying co1 show variable chlorophyll content. Co2.1 is very similar to Wt.



Supplemental Figure 3: Semi-quantitative PCR on *Cg*Glk1 and *At*Actin7 in three complemented co1 and co2 lines. Co1.1 and Co2.1 show the most intense *Cg*Glk1 signal and were used in subsequent experiments.

Supplemental dataset 1: RPKM values as reads mapped to AGIs

Supplemental dataset 2: List with oligonucleotides

Supplemental dataset 3: differentially expressed genes in glk2 (expression >20 RPKM)

(Supplemental datasets are stored on the attached disc)

References

Allahverdiyeva, Y., Mamedov, F., Holmström, M., Nurmi, M., Lundin, B., Styring, S., Spetea, C., and Aro, E.-M. (2009). Comparison of the electron transport properties of the< i> psbo1</i> and< i> psbo2</i> mutants of< i> Arabidopsis thaliana</i>. Biochimica et Biophysica Acta (BBA)-Bioenergetics **1787,** 1230-1237.

Barber, J., and Andersson, B. (1992). Too much of a good thing: light can be bad for photosynthesis. Trends in Biochemical Sciences **17,** 61-66.

Barker, M.S., Vogel, H., and Schranz, M.E. (2009). Paleopolyploidy in the Brassicales: Analyses of the Cleome Transcriptome Elucidate the History of Genome Duplications in Arabidopsis and Other Brassicales. Genome Biology and Evolution **1,** 391-399.

Bravo-Garcia, A., Yasumura, Y., and Langdale, J.A. (2009). Specialization of the Golden2-like regulatory pathway during land plant evolution. The New phytologist **183,** 133-141.

Brown, N.J., Parsley, K., and Hibberd, J.M. (2005). The futured C-4 research - Maize, Flaveria or Cleome? Trends Plant Sci. **10,** 215-221.

Buick, R. (1992). The antiquity of oxygenic photosynthesis: evidence from stromatolites in sulphate-deficient Archaean lakes. Science **255,** 74-77.

Chalker-Scott, L. (1999). Environmental Significance of Anthocyanins in Plant Stress Responses. Photochemistry and Photobiology **70,** 1-9.

Cheng, S., van den Bergh, E., Zeng, P., Zhong, X., Xu, J., Liu, X., Hofberger, J., de Bruijn, S., Bhide, A.S., Kuelahoglu, C., Bian, C., Chen, J., Fan, G., Kaufmann, K., Hall, J.C., Becker, A., Brautigam, A., Weber, A.P., Shi, C., Zheng, Z., Li, W., Lv, M., Tao, Y., Wang, J., Zou, H., Quan, Z., Hibberd, J.M., Zhang, G., Zhu, X.G., Xu, X., and Schranz, M.E. (2013). The Tarenaya hassleriana genome provides insight into reproductive trait and genome evolution of crucifers. The Plant cell **25,** 2813-2830.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana. The Plant Journal **16,** 735-743.

Curtis, M.D., and Grossniklaus, U. (2003). A Gateway Cloning Vector Set for High-Throughput Functional Analysis of Genes in Planta. Plant Physiol. **133,** 462-469.

Deblaere, R., Bytebier, B., De Greve, H., Deboeck, F., Schell, J., Van Montagu, M., and Leemans, J. (1985). Efficient octopine Ti plasmid-derived vectors for Agrobacterium-mediated gene transfer to plants. Nucleic Acids Res. **13,** 4777-4788.

Engelmann, S., Zogel, C., Koczor, M., Schlue, U., Streubel, M., and Westhoff, P. (2008). Evolution of the C4 phosphoenolpyruvate carboxylase promoter of the C4 species Flaveria trinervia: the role of the proximal promoter region. BMC Plant Biology **8,** 4.

Farquhar, J., Zerkle, A., and Bekker, A. (2011). Geological constraints on the origin of oxygenic photosynthesis. Photosynthesis Research **107,** 11-36.

Ferreira, K.N., Iverson, T.M., Maghlaoui, K., Barber, J., and Iwata, S. (2004). Architecture of the Photosynthetic Oxygen-Evolving Center. Science **303,** 1831-1838.

Fiehn, O. (2007). Validated high quality automated metabolome analysis of Arabidopsis thaliana leaf disks. In Concepts in Plant Metabolomics (Springer), pp. 1-18.

Fitter, D.W., Martin, D.J., Copley, M.J., Scotland, R.W., and Langdale, J.A. (2002). GLK gene pairs regulate chloroplast development in diverse plant species. The Plant journal : for cell and molecular biology **31,** 713-727.

García-Cerdán, J.G., Sveshnikov, D., Dewez, D., Jansson, S., Funk, C., and Schröder, W.P. (2009). Antisense Inhibition of the PsbX Protein Affects PSII Integrity in the Higher Plant Arabidopsis thaliana. Plant Cell Physiol. **50,** 191-202.

Genty, B., Briantais, J.-M., and Baker, N.R. (1989). The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. Biochimica et Biophysica Acta (BBA) - General Subjects **990,** 87-92.

Gould, K.S., McKelvie, J., and Markham, K.R. (2002). Do anthocyanins function as antioxidants in leaves? Imaging of H2O2 in red and green leaves after mechanical injury. Plant, Cell & Environment **25,** 1261-1269.

Gowik, U., and Westhoff, P. (2011). The Path from C3 to C4 Photosynthesis. Plant Physiol. **155,** 56-63.

Hall, L.N., Rossini, L., Cribb, L., and Langdale, J.A. (1998). GOLDEN 2: A Novel Transcriptional Regulator of Cellular Differentiation in the Maize Leaf. The Plant Cell Online **10,** 925-936.

Hsia, C.C., and McGinnis, W. (2003). Evolution of transcription factor function. Current Opinion in Genetics & Development **13,** 199-206.

Ishihara, S., Takabayashi, A., Ido, K., Endo, T., Ifuku, K., and Sato, F. (2007). Distinct Functions for the Two PsbP-Like Proteins PPL1 and PPL2 in the Chloroplast Thylakoid Lumen of Arabidopsis. Plant Physiol. **145,** 668-679.

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

Lee, D.Y., and Fiehn, O. (2008). High quality metabolomic data for Chlamydomonas reinhardtii. Plant methods **4,** 7.

Majeran, W., and van Wijk, K.J. (2009). Cell-type-specific differentiation of chloroplasts in C4 plants. Trends Plant Sci. **14,** 100-109.

Majeran, W., Cai, Y., Sun, Q., and van Wijk, K.J. (2005). Functional Differentiation of Bundle Sheath and Mesophyll Maize Chloroplasts Determined by Comparative Proteomics. The Plant Cell Online **17,** 3111-3140.

Majeran, W., Zybailov, B., Ytterberg, A.J., Dunsmore, J., Sun, Q., and van Wijk, K.J. (2008). Consequences of C4 Differentiation for Chloroplast Membrane Proteomes in Maize Mesophyll and Bundle Sheath Cells. Molecular & Cellular Proteomics **7,** 1609-1638.

Michelet, L., Zaffagnini, M., Morisse, S., Sparla, F., Pérez-Pérez, M.E., Francia, F., Danon, A., Marchand, C.H., Fermani, S., and Trost, P. Redox regulation of the Calvin–Benson cycle: something old, something new.

Moore, R., and Black, C.C. (1979). Nitrogen Assimilation Pathways in Leaf Mesophyll and Bundle Sheath Cells of C4 Photosynthesis Plants Formulated from Comparative Studies with Digitaria sanguinalis (L.) Scop. Plant Physiol. **64,** 309-313.

Murakami, R., Ifuku, K., Takabayashi, A., Shikanai, T., Endo, T., and Sato, F. (2005). Functional dissection of two Arabidopsis PsbO proteins. FEBS Journal **272,** 2165-2175.

Porra, R. (2002). The chequered history of the development and use of simultaneous equations for the accurate determination of chlorophylls a and b. Photosynthesis Research **73,** 149-156.

Raven, J.A., Cockell, C.S., and De La Rocha, C.L. (2008). The evolution of inorganic carbon concentrating mechanisms in photosynthesis.

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

Romeis, T., Ludwig, A.A., Martin, R., and Jones, J.D.G. (2001). Calcium-dependent protein kinases play an essential role in a plant defence response. The EMBO Journal **20,** 5556-5567.

Rossini, L., Cribb, L., Martin, D.J., and Langdale, J.A. (2001). The Maize Golden2 Gene Defines a Novel Class of Transcriptional Regulators in Plants. The Plant Cell Online **13,** 1231-1244.

Sheen, J. (1991). Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes. The Plant Cell Online **3,** 225-245.

Suorsa, M., Sirpiö, S., Allahverdiyeva, Y., Paakkarinen, V., Mamedov, F., Styring, S., and Aro, E.-M. (2006). PsbR, a Missing Link in the Assembly of the Oxygen-evolving Complex of Plant Photosystem II. Journal of Biological Chemistry **281,** 145-150.

Tanaka, R., and Tanaka, A. (2007). Tetrapyrrole Biosynthesis in Higher Plants. Annual Review of Plant Biology **58,** 321-346.

Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, 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 : for cell and molecular biology **37,** 914-939.

Thornber, J.P. (1975). Chlorophyll-proteins: light-harvesting and reaction center components of plants. Annual Review of Plant Physiology **26,** 127-158.

Tuch, B.B., Li, H., and Johnson, A.D. (2008). Evolution of Eukaryotic Transcription Circuits. Science **319,** 1797-1799.

Usadel, B., Nagel, A., Steinhauser, D., Gibon, Y., Blasing, 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.

Vitha, S. (2007). Histochemical localization of β-glucuronidase (GUS) reporter activity in plant tissues. Microscopy and Imaging Center.

Wallsgrove, R.M., Keys, A.J., Lea, P.J., and Miflin, B.J. (1983). Photosynthesis, photorespiration and nitrogen metabolism. Plant, Cell & Environment **6,** 301-309.

Wang, P., Fouracre, J., Kelly, S., Karki, S., Gowik, U., Aubry, S., Shaw, M.K., Westhoff, P., Slamet-Loedin, I.H., Quick, W.P., Hibberd, J.M., and Langdale, J.A. (2013). Evolution of GOLDEN2-LIKE gene function in C(3) and C(4) plants. Planta **237,** 481-495.

Waters, M.T., Moylan, E.C., and Langdale, J.A. (2008). GLK transcription factors regulate chloroplast development in a cell-autonomous manner. The Plant Journal **56,** 432-444.

Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J., and Langdale, J.A. (2009). GLK Transcription Factors Coordinate Expression of the Photosynthetic Apparatus in Arabidopsis. The Plant Cell Online **21,** 1109-1128.

Yasumura, Y., Moylan, E.C., and Langdale, J.A. (2005). A Conserved Transcription Factor Mediates Nuclear Control of Organelle Biogenesis in Anciently Diverged Land Plants. The Plant Cell Online **17,** 1894-1907.

Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., and Liu, X. (2014). Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. PloS one **9,** e78644.

Manuscript-draft 4

A specific, cholesterol-binding Lipid Transfer Protein is highly abundant in the C_4 plant *Cleome gynandra*

Manuel Sommer*[,1], Andrea Bräutigam[1], Elia Stahl[2], Jürgen Zeier[2], Andreas PM Weber[a,1]

[1] Plant Biochemistry, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany

[2] Molecular Plant Ecophysiol, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany

* first author

[a] corresponding author

Abstract

Plant lipid transfer proteins (LTPs) build large gene families and facilitate a large variety of lipid transport processes *in vitro*. Although substrate specific LTPs were characterized in yeast, plant LTPs are non-specific (nsLTPs) towards their substrate, as indicated by equal transport rates of different phospholipid species. Here, we functionally characterize an LTP that is highly abundant in the C_4 species *Cleome gynandra*. We localized *Cg*LTP to the extracellular space of bundle sheath cells with GUS staining and fluorescence tagging. Heterologously expressed *Cg*LTP exclusively bound to cholesterol in a protein-lipid binding assay, while no binding was detected for other plant sterols, wax precursors, fatty acids or phospholipids. The content of free cholesterol in *C. gynandra* was lower than in a closely related C_3 species. We conclude that *Cg*LTP is a secreted cholesterol binding protein that might be involved in the development of C_4 bundle sheath tissues.

Introduction

Plant lipids are synthesized in chloroplasts and the endoplasmatic reticulum(Ohlrogge and Browse, 1995). It is unknown, how lipids are transferred from these biosynthesis sites to other organelles or the extracellular space. *In vitro*, lipid transfer proteins (LTPs) have the ability to exchange phospholipids between membranes, but a predominant localization of LTPs to the extracellular space makes a role in intracellular lipid metabolism unlikely (Kader, 1975; Sterk et al., 1991; Coutos-Thevenot et al., 1993). LTPs were identified in bacteria, fungi, plants and mammals (Wirtz, 1991), but plant LTPs share no sequence homology with mammalian LTPs (Thoma et al., 1993). In plants, LTPs form large gene families that can account for up to 4% of total soluble proteins (Kader, 1996; Arondel et al., 2000; Boutrot et al., 2008). LTPs are small, basic proteins (~9 kDa). The substrates that were shown to be transferred by LTPs include phospholipids, glycosylated lipids, fatty acids and sterols (Arondel and Kader, 1990; Yamada, 1992). Many LTPs were shown to transfer various lipids between membranes and are therefore referred to as non-specific LTPs (nsLTPs) (Kader et al., 1984; Lascombe et al., 2008). Specific binding of lipids is displayed by a wide variety of non-LTP proteins, but knowledge about specific binding of lipid species by LTPs is scarce (Veerkamp et al., 1991). An LTP in yeast was found to bind to phosphatidylinositol specifically (Sha and Luo, 1999). However, to our knowledge substrate exclusivity was never determined in a plant LTP. The biological role of plant LTPs is not resolved, though there are some indications that LTPs are involved in modelling of extracellular structures and stress response (Douliez et al., 2000).

When the transcriptome of two closely related species of the family of *Cleomaceae* was compared, an LTP was found highly up-regulated in *Cleome gynandra*, a plant performing C_4 photosynthesis as compared to its C_3 relative, *Tarenaya hassleriana* (Külahoglu et al., 2014). The expression level of the LTP from C. gynandra (*Cg*LTP) was dynamic and increased during leaf development. C_4

photosynthesis is a complex trait that requires adaptation of the photosynthetic apparatus, leaf anatomy and it compartmentalizes whole metabolic pathways between two distinct cell types: the central bundle sheath and the peripheral mesophyll tissue (Sage et al., 2012). Subsequent compartmentalization of central metabolic pathways was essential for functional C_4 photosynthesis. However, scarce knowledge exists about anatomical adaptations to the special needs of C_4 photosynthesis.

In this study, we functionally characterized *Cg*LTP to understand its role in C_4 photosynthesis. Therefore, we compared the lipidome of *C. gynandra* and *T. hassleriana* and the binding specificity of *Cg*LTP towards various plant lipids. Furthermore, we analyzed its localization by heterologous expression in the model plant *Arabidopsis thaliana*.

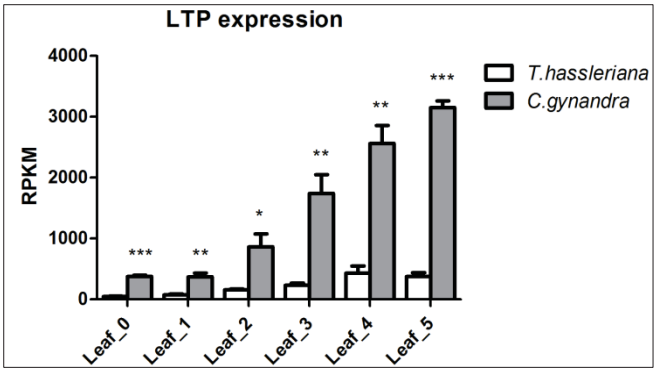


Figure 1: Expression levels of *Cg*LTP and its closest related LTP in *T. hassleriana* (accession number 24262). Expression is normalized to reads per kilobase and million (RPKM). Samples represent developing leaf stages. Leaf_0 is the youngest stage and Leaf_5 the oldest. Asterisks indicate significance between *C. gynandra* and *T. hassleriana* with ***=p<0.001, **=p<0.01 and *=p<0.05. Original data from: (Külahoglu 2014)

Material and Methods

Chemicals and reagents

Chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Reagents for the ß-Glucuronidase assay were purchased from Duchefa (Haarlem, Netherlands).

Cultivation of plant species

C. gynandra and *T. hassleriana* plants were grown on soil in a greenhouse. The light period was enhanced to 16 hours per day by artificial illumination with 150 µE, growth temperature was 24°C and relative humidity was 70%.

A. thaliana seeds were vernalized and subsequently germinated in petri dishes on solid MS media in a Percival CU-36L5/D incubator (Percival, Perry, IA). Twelve days after germination, plants were transferred to 9 cm pots containing soil and were grown in Percival CU-36L5 incubators (Percival, Perry, IA). In both incubators, plants were illuminated 12 hours a day with a light intensity of 100 µE at 22°C during the day and 18°C during the night.

All plants were watered with 0.2% Wuxal 8+8+6 fertilizer (Bayer, Leverkusen, Germany) every seven days, additional water was applied on demand.

When leaf extract was required for an experiment, three leaves of three plants were cut and frozen in liquid nitrogen immediately (< 1s). Frozen leaves were then ground to a fine powder under constant supply with liquid nitrogen using mortar and pestle. The powder was stored at -80°C until used.

Phylogenetic analysis

For phylogenetic analysis, the sequence of *CgLTP* was extracted from RNA sequencing data and Blast analysis was performed on the TAIR9 coding sequence dataset (www.arabidopsis.org) and the *T. hassleriana* genome (Cheng et al., 2013). Sequences that were similar to *Cg*LTP with e < 10^{-10} were included in phylogenetic analysis. Phylogenetic analysis was carried out on phylogeny.fr (www.phylogeny.lirmm.fr) using the MUSCLE alignment algorithm, Gblocks curation and the PhyML phylogeny software. Multiple sequence alignment was done with ClustalW2 (www.ebi.ac.uk/Tools/msa/clustalw2) and edited in Microsoft Word (Microsoft, Redmond, WA).

Semi-quantitative RTPCR on LTP genes of *C. gynandra* and *T. hassleriana*

Template cDNA was generated from leaf powder of *C. gynandra* and *T. hassleriana*. Total RNA was extracted by an RNase all protocol (Chomczynski and Sacchi, 1987) and cDNA first strand synthesis was done with SuperScript II reverse transcriptase (Invitrogen, Carlsbad, CA) according to the producer's instructions. Semi-quantitative reverse transcription (semi-q RT) polymerase chain reaction (PCR) was done on cDNA using GoTaq polymerase (New England Biolabs, Frankfurt a.M., Germany) following the manufacturer's instructions. Amplification primers for LTP quantification were P1/P2 and P3/P4 for *C. gynandra* and *T. hassleriana*, respectively. Actin7 primers were designed on the most similar nucleotide sequence to Actin7 from *A. thaliana* in the *C. gynandra* transcriptome (Külahoglu 2014, P5/P6) and the *T. hassleriana* genome (P7/P8) (Cheng et al., 2013). Bands were quantified with the ImageJ software (ImageJ.net).

Thin layer chromatography on plant extract

Lipids were extracted from 50 mg leaf powder by adding 1 mL methanol:chloroform (2:1). After 20 min incubation, 0.3 mL chloroform and 0.6 mL water were added. Samples were mixed, separated by centrifugation and the organic phase was dried under a nitrogen stream. Samples were re-solved in 200 µL methanol:chloroform (1:1).

Thin layer chromatography (TLC) was carried out on silica gel plates (HPLTC Silica gel 60, Millipore, Billerica, MA) using three different solvents to separate the lipid classes. Neutral lipids were separated with hexane:diethyl ether:acetic acid (90:15:2) (Juguelin et al., 1986), phospholipids were separated with methyl acetate:n-propanol:chloroform:methanol:0.25% aqueous KCl (25:25:25:10:9) (Vitello and Zanetta, 1978) and glycolipids were separated with diisobutyl ketone:acetic acid:water (80:50:10) (Lepage, 1964). Thin layer chromatography was stopped when the solvent reached the top of the

plate. Lipids were visualized by incubating the dried TLC plates for 5 seconds in 8% H_3PO_4 that contained 3% $CuSO_4$. Subsequently, TLC plates were charred in an oven at 130°C for 10 minutes as summarized by Bitman (Bitman and Wood, 1982).

Expression of ß – glucuronidase in reporter lines and staining assay

The *Cg*LTP promoter was amplified via PCR (P9/P10) and cloned into pCambia1381 (Cambia, Canberra, Australia) over SpeI/PstI restriction sites. Agrobacterium tumefaciens strain GV2260 (Deblaere et al., 1985) was transformed with the resulting plasmid and flowering *A. thaliana* plants were infected with the bacteria by floral dip (Clough and Bent, 1998). The offspring of infected plants was screened on solid MS media containing Hygromycin (33 µg/mL). 19 plants that were able to grow on selective media, were selfed and the offspring was screened in a ß – Glucuronidase assay (Vitha, 2007). Blue staining was consistently found close to veins in all lines.

Expression of a yellow fluorescent protein (Venus) in reporter lines and microscopy

The *Cg*LTP gene (amplified via PCR with P11/P12) was integrated into the pUBQ10 promoter-driven (Norris et al., 1993; Krebs et al., 2012) pHygII-UT-c-term-Venus vector (Walter et al., 2004; Nagaya et al., 2010) (kindly provided by Jörg Kudla's laboratory) using the XmaI and XbaI restriction site. Agrobacterium tumefaciens strain GV2260 (Deblaere et al., 1985) was transformed with the resulting plasmid and *Nicotiana benthamiana* leaves were infiltrated with a suspension of bacterial cells (Romeis et al., 2001). Two days after infiltration, the epidermal tissue was peeled off and observed under a laser scanning microscope (Zeiss LSM510, Carl Zeiss, Jena, Germany) with a YFP filter (excitation 504 nm; emission 542 nm). A digital merge of the pictures was assembled with the Zeiss Zen 2011 software (Carl Zeiss, Jena, Germany). Protoplasts were extracted from equally transformed leaves after Sheen (Sheen, 1991) and were observed under a microscope like the epidermal peels.

Lipid-protein binding assays with *Cg*LTP

*Cg*LTP was cloned into pIVEX1.4 via conventional cloning (P13/P14) (5 Prime, Gaithersburg, MD) and heterologously expressed with the RTS 100 wheat germ CECF kit (5 Prime, Gaithersburg, MD) according to the manufacturer's instructions. Lipids were solved in methanol:chloroform (1:1). For the binding assay, a nitrocellulose membrane was spotted with 25 µg (assay 1) or 10 and 50 µg (assay 2) of all lipids. For assay 1, membranes were dried and a protein-lipid overlay assay (Dowler et al., 2002) was performed with a 1:200 dilution of unpurified *Cg*LTP in wheat germ extract and a 1:5000 dilution of a rabbit anti HIS HRP conjugate antibody (Miltenyi, Bergisch-Gladbach, Germany). For assay 2, all buffers were prepared without detergent. HRP staining was done with the Immobilon Chemiluminescent HRP substrate (Millipore, Billerica, MA). Stains were visualized with a LAS-4000 mini (GE Health Care, Piscataway, NJ) and the associated LAS software.

Detection of free sterols via GC/MS

Sterols were extracted from 50 mg leaf powder via vapor phase extraction (Mishina and Zeier, 2006) and subsequently quantified via GC/MS analysis (Griebel and Zeier, 2010) on a gas chromatograph (GC 7890 A, Agilent, Santa Clara, CA) equipped with a fused silica capillary column (HP-1 MS 30m x 0.25 mm, Agilent, Santa Clara, CA). Mass spectra were recorded with a combined 5975 C mass spectrometric detector (Agilent, Santa Clara, CA) in the electron ionization mode. Chromatograms of specific mas-charge ratios were analyzed for sterol quantification (cholesterol: m/z 329, ß-sitosterol: m/z 486, campesterol: m/z 382, stigmasterol: m/z 380, ergosterol: m/z 468). Peak areas were quantified relatively by normalization to the internal standard ergosterol.

Mevastatin treatment and vein quantification

Emerging leaves of twenty-one day old *C. gynandra* and *T. hassleriana* plants were completely covered with 10 mM Mevastatin (treatment) or with water (control) once a day for a period of five days. The solutions were spread across the leaf surface with a fine brush. After seven days, three 3-4 cm long treated leaves were harvested and bleached in ethanol:acetic acid (3:1) for 24 hours. Bleached leaves were stained with 0.5% safranine in ethanol for 1 hour. Leaves were washed with ethanol three times for 10 minutes and observed with an inverted microscope (Eclipse Ti-U, Nikon, Tokyo, Japan) and the ProgRes software (Jenoptik, Jena, Germany). Vein density was quantified by dividing the sum of all vein lengths by the area they were counted in. Four pictures from different areas of 6 mm^2 were taken and quantified per leaf.

Results

*Cg*LTP is highly abundant in the C_4 plant *C. gynandra*

In this study, we characterized the function of a lipid transfer protein from the C_4 plant *C. gynandra*. We derived the coding sequence from RNA sequencing analysis on *C. gynandra* and conducted phylogenetic analysis with *Cg*LTP and its best matches from Blast analysis in two closely related species. Among the 16 best Blast hits in *A. thaliana* we found 13 genes coding for LTPs. Furthermore, nine genes that are similar to *Cg*LTP according to Blast were added from the *T. hassleriana* genome (Cheng et al., 2013). Phylogenetic analysis showed that *Cg*LTP is most closely related to annotation 24262 of its sister species *T. hassleriana*. The closest gene from *A. thaliana* is At1G12090, which is annotated as an Extensin-Like Protein.

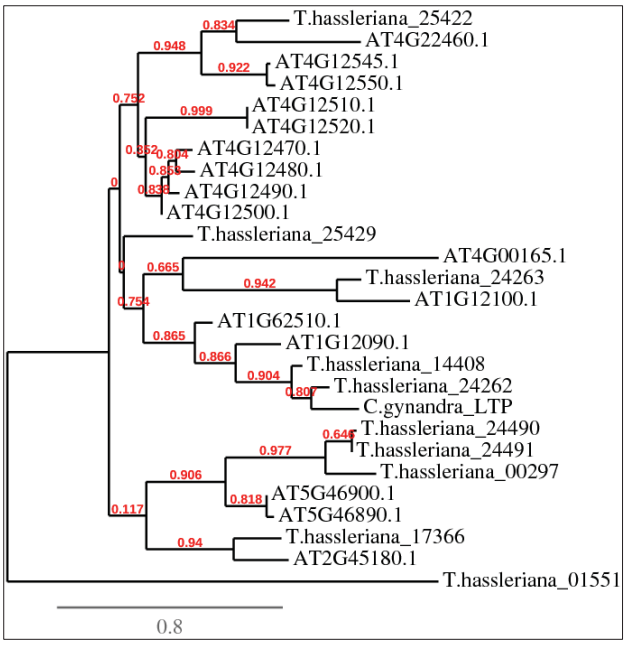


Figure 2: Phylogenetic comparison of nucleotide sequences of *Cg*LTP and its closest relative genes from *A. thaliana* and *T. hassleriana* done with an online tool at phylogeny.lirmm.fr. Red numbers indicate branch support after 1000 bootstraps, length of branches is proportional to the number of nucleotide exchanges per site.

We tested conservation of the amino acid sequence of *Cg*LTP and its two best hits from *A. thaliana* and *T. hassleriana*. Amino acid identity between *Cg*LTP and its closest relative in *T. hassleriana* was 89%, while this value was 75% between *Cg*LTP and its best match in *A. thaliana*. We found eight cysteine residues in an assembly that is characteristic for LTPs in *Cg*LTP and its close relatives (Yeats and Rose, 2008). The protein is 14 kDa in size and is therefore larger than other LTPs, which is the result of a long N-terminal target peptide that is predicted with high reliability (>0.8) by TargetP. Beginning with the conserved eight cysteine motif, *Cg*LTP contains 84 amino acids with a molecular mass of 8.6 kDa. It is thus between family 1 and family 2 LTPs, which are 9 kDa and 7 kDa in size, respectively. *Cg*LTP furthermore shares only few conserved sites of family 1 or 2 LTPs besides the conserved cysteine residues (Yeats and Rose, 2008).

```
AT1G62510.1             MASRTTKSLALFLILNFLFFTTISACGNCGCPSPKPKHKPSPSPKPKPNPKPKPTPHPSP
AT1G12090.1             ---MASSSIALFLALNLLFFTTISACGSCTPC-----GGGCPSPKP--KPTPK--PTPSP
T.hassleriana_14408     MASKTKISLAIFLLLNLLFFTLTSACNTCSPC-----GGGCPSPKP--KPRPNP----NP
T.hassleriana_24262     MASKTTSSLAIFLLLNLLFFTLTSACNSCSPC-----GSGCPSPKP--KPKPKPKPSPSP
C.gynandra_LTP          MASKSSISLAIFLLLNLLFFTLTTACNSCNPC-----GNGCPSPKP--KPNPNPKPTPSP
                            :. *:*:** **:****  :** .*           .*****  :* *:     .*

AT1G62510.1             SPAIAKCPRDALKLGVCANVLNGLLNVTLGKPPVEPCCTLIQGLADLEAAACLCTALKAN
AT1G12090.1             SSGSSKCPKDTLKLGVCANVLNGLLDLTLGKPPVEPCCSLIQGLADVEAAVCLCTALKAN
T.hassleriana_14408     SPAVAKCPKDTLKLGVCANVLNGLLNLQLGQPPVTPCCSLIQGLADVEAAVCLCTALKAN
T.hassleriana_24262     SPAQAKCPKDALKLGVCANVLNGLLNVTLGQPPVEPCCTLIQGLADVEAAACLCTALKAN
C.gynandra_LTP          NPSEAKCPKDTLKLGVCANVLNGLLNVTLGQPPVTPCCTLLQGLADVEAAACLCTALKAN
                        . . :***:*:***************:: **:*** ***:*:*****:***.*********

AT1G62510.1             ILGINLNIPLSLSLLLNVCSKKVPRGFQC*-
AT1G12090.1             ILGINLNLPISLSLLLNVCSKQLPPGFQC*-
T.hassleriana_14408     ILGINLNLPISLSLLLNVCSKKVPSGFQCE*
T.hassleriana_24262     ILGINLNLPISLSLLLNVCSKKVPPGFQCE*
C.gynandra_LTP          ILGINLNLPISLSLLLNVCSKKVPSGFQCE*
                        *******:*:***********::* ****
```

Figure 3: Multiple sequence alignment from Clustal Omega. *Cg*LTP and its two closest relatives from both *A. thaliana* and *T. hassleriana* were included. The conserved eight cysteine residues are highlighted in grey. Conserved sites are indicated by asterisks, similar sites are indicated by one (less similar) or two (more similar) dots.

To complement RNA sequencing information from previous studies, we analyzed CgLTP and *Th*LTP (accession 24262) expression levels in mature leaves of *C. gynandra* and *T. hassleriana*, respectively. Therefore we performed semi-q RTPCR on cDNA from leaf extract. Expression of *Cg*LTP was found significantly increased as compared to its closest relative in *T. hassleriana*.

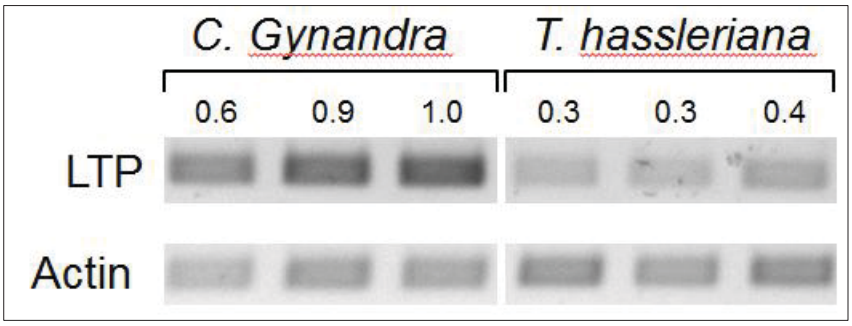


Figure 4: Semi-q RTPCR on leaf cDNA from *C. gynandra* and *T. hassleriana*. Numbers represent signal strength of the LTP signal as calculated by the ImageJ software. Actin7 signals serve as internal control

The lipid composition of C. gynandra and T. hassleriana is largely similar

We analyzed the lipid composition of *C. gynandra* and *T.hassleriana* to discover putative consequences of an increased expression of *Cg*LTP. Therefore, we extracted total lipids from leaf extract and performed thin layer chromatography with three different solvents. Lipid composition of both species was almost identical. The main lipid species in both species were MGDG and DGDG. Phospholipids were found minor components but also equally abundant in both species. When neutral

lipids were separated, a single fatty alcohol species was visualized in *T. hassleriana* that was lacking in *C. gynandra*, while another fatty alcohol species was more prominent in the C_4 plant.

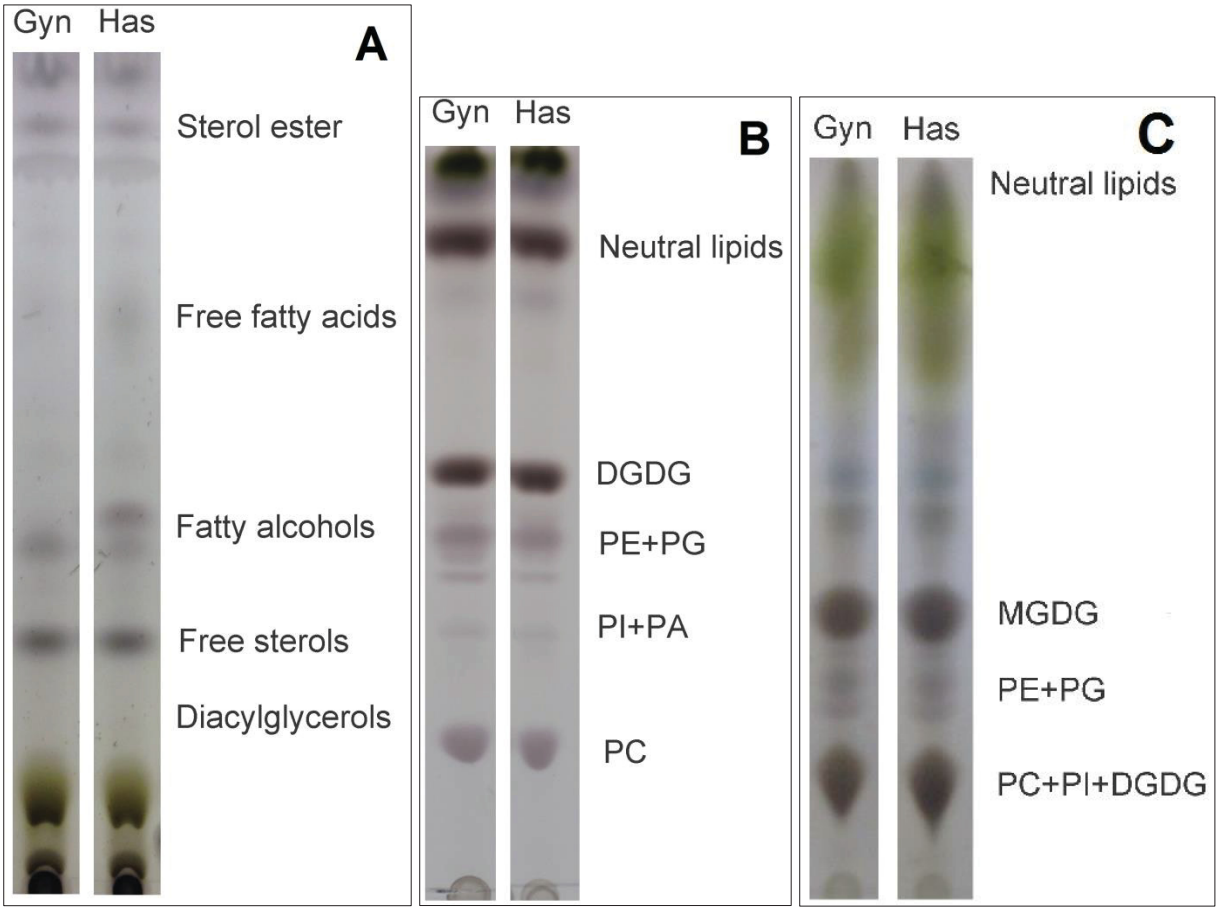


Figure 5: Thin layer chromatograms of lipid extracts from *C. gynandra* and *T. hassleriana* leaves. Lipid standards (1-docosanol, cholesterol, PE, PC, arachidic acid) were used to determine lipid classes in the separated samples (bands not shown). Neutral lipids (A), phospholipids (B) and glycolipids (C) were separated with different solvents. Gyn=*C. gynandra*, Has=*T. hassleriana*.

CgLTP localizes to the cell periphery of cells of the vascular bundle

To determine the localization of *Cg*LTP, we transformed *A. thaliana* plants with a *Cg*LTPpromoter:GUS fusion protein and visualized the protein in a biochemical assay. GUS-mediated staining was found exclusively in tissues that were close to leaf veins, while no staining was observed in mesophyll tissue or root. To elucidate the subcellular localization of *Cg*LTP, we transiently transformed *Nicotiana benthamiana* plants with a CgLTP:Venus fusion protein. We visualized the construct by fluorescence microscopy and found a signal in the periphery of transformed epidermal cells. This signal was not present in protoplasts that were extracted from leaf tissue prior to microscopy (data not shown). A fluorescence signal was also seen in scattered organelles that were

identified as chloroplasts. This fluorescence signal, however, was seen in all cells independent of the peripheral signal.

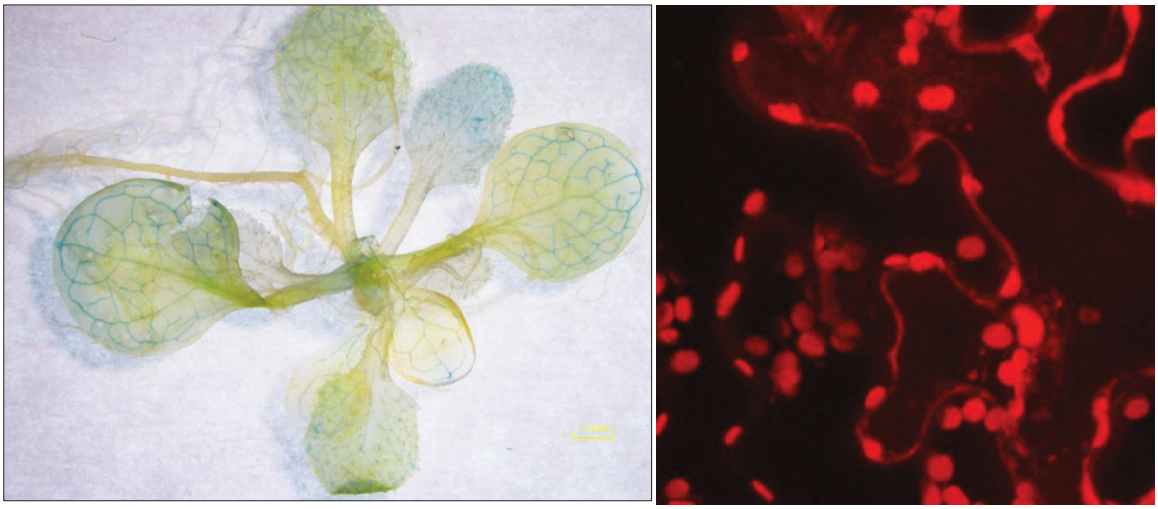


Figure 6: GUS staining of *A. thaliana* seedling carrying a GUS gene driven by the *Cg*LTP promoter.

Figure 7: YFP fluorescence signal of an epidermis peel of an *N. benthamiana* leaf. Leaves were transiently transformed with a *Cg*LTP:Venus fusion gene driven by the pUBQ10 promoter 48 hours before.

*Cg*LTP binds cholesterol specifically

To test whether *Cg*LTP can bind to lipid species, we expressed the protein in wheat germ extract and conducted a protein-lipid binding assay on a nitrocellulose membrane (fat-western, assay 1). We tested for binding to fatty acids and sterols on the membrane. Upon visualization, a dark stain appeared where cholesterol was spotted, while all other spots remained blank.

We repeated the assay with an increased number of putative substrates and washed the nitrocellulose membrane without any detergents, to be able to detect weak interactions of *Cg*LTP with lipids (assay 2). We applied two different concentrations (10 and 50 µg) of all lipids to determine quantitative differences in substrate binding of *Cg*LTP. Dark stains appeared only at spots where cholesterol was applied. In addition, we found a faint stain at the 50 µg spot of Campesterol. This signal was significantly weaker than the cholesterol signal, when 10 µg were applied. Other spots were stained diffusely but showed no distinct circular stain as seen where cholesterol was applied. Among unstained lipids was the fatty alcohol 1-Docosanol.

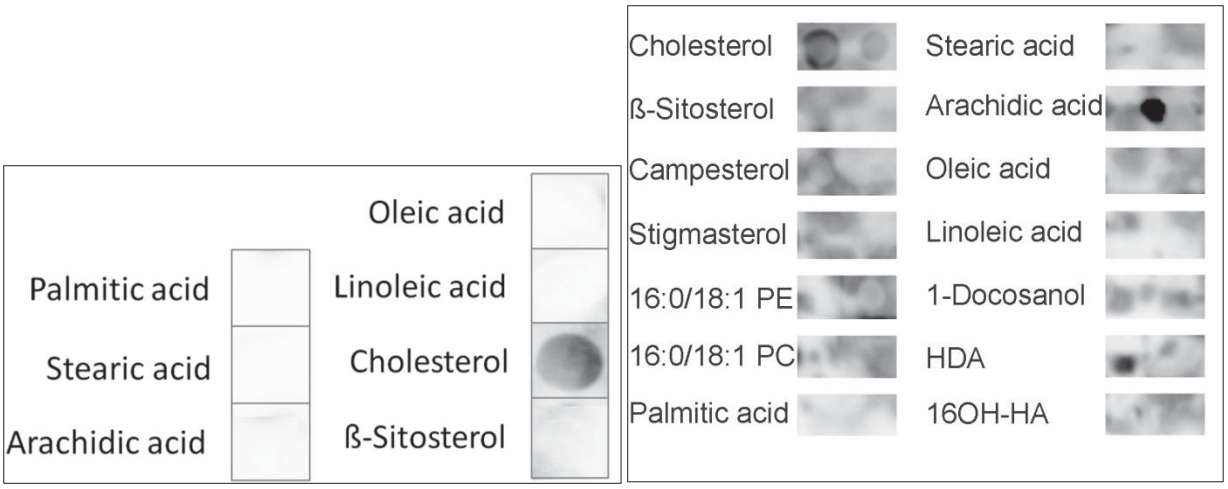


Figure 8: Lipid-protein binding assay (assay 1) with heterologously expressed *Cg*LTP on fatty acid and sterol species. 25 ug of each lipid were spotted on the nitrocellulose membrane.

Figure 9: Lipid-protein binding assay with extended substrates and without the use of detergents (assay2). 50 µg (left spot) and 10 µg (right spot) were applied to the membrane.

C. gynandra contains less cholesterol than its close relative C_3 species, *T. hassleriana*

Next, we measured the content of free sterols in two pairs of closely related C_3 and C_4 plants. The most prominent sterol was ß-Sitosterol in all four species. The three less abundant sterols, cholesterol, stigmasterol and campesterol, contributed 27% to the total free sterol content. *T. hassleriana* contained 2.2 times more ß-sitosterol than *C. gynandra*. A similar trend was not observed between *F. bidentis* and *F. robusta*. Comparative analysis showed that the content of all four sterols was significantly different between *C. gynandra* and *T. hassleriana*. The pair of *Flaveria* species showed a different abundance for only two of four sterols. When we calculated foldchanges between C_4 and C_3 plants, we found a common trend for stigmasterol, which is more abundant in both C_4 plants, while the other sterols do not show a clear trend between C_3 and C_4 species.

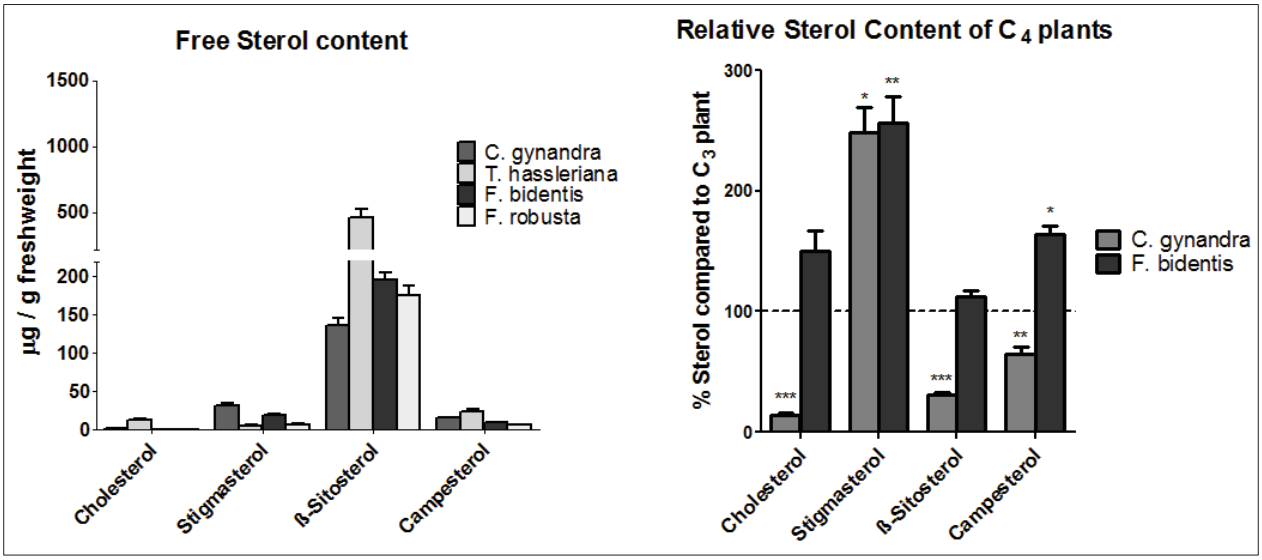


Figure 10: Content of four sterols in two pairs of closely related C_3 and C_4 species (A). Relative sterol content in the C_4 plants *C. gynandra* and *F. bidentis* compared to their closely related C_3 species, *T. hassleriana* and *F. robusta* (B). Dashed line indicates sterol contents of the C_3 plant (100%), asterisks indicate significance between C_3 and C_4 species, where ***=p<0.001, **=p<0.01, *=p<0.05 in a Student's t-test.

A disturbed sterol metabolism does not affect vein formation

To test whether cholesterol content has an effect on vein patterning of a C_4 leaf, we treated developing leaves of *C. gynandra* with the cholesterol biosynthesis inhibitor Mevastatin. We counted veins after five days of treatment and found a mild decrease of vein density in Mevastatin-treated leaves that was found not significant.

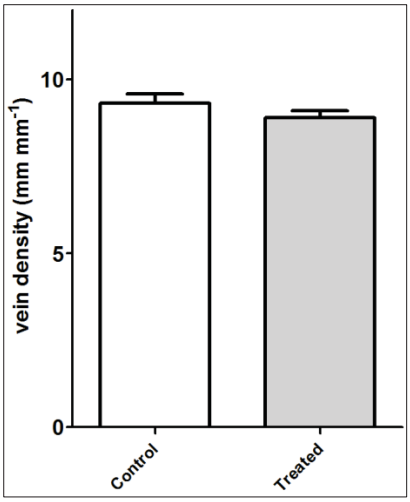


Figure 11: Average vein density after Mevastatin treatment and untreated control

Discussion

In this study we functionally characterize a lipid transfer protein from the C_4 plant *C. gynandra*. The protein binds to cholesterol, while it does not bind to other sterols or plant lipids. We therefore term it a cholesterol binding protein.

In a comparative transcriptome study, a lipid transfer protein was highly up-regulated in the C_4 plant *C. gynandra* as compared to its close relative C_3 species *T. hassleriana* (Külahoglu et al., 2014). We extracted the coding sequence from RNA sequencing information and perfomed Blast analysis to find close homologs in *A. thaliana* and *T. hassleriana*, where we found genes that shared 75% and 89% of their amino acid sequence with *Cg*LTP. High amino acid similarity was expected, since *C. gynandra* and *T. hassleriana* are sister species in the family *Cleomaceae*, which belongs to the order of *Brassicales*, like *A. thaliana* (Inda et al., 2008). Furthermore, lipid transfer protein sequences were shown to be conserved between plant species (Tchang et al., 1988). When we analyzed the amino

acid sequence of *Cg*LTP and its closest homologs in *A. thaliana* and *T. hassleriana*, we found eight cysteine residues in a formation that is common for all LTPs ($CX_nCX_nCCX_nCXCX_nCX_nC$) and responsible for the protein structure (Kader, 1996; Douliez et al., 2000). However, other features of the *Cg*LTP amino acid sequence are unusual for plant LTPs. The predicted cleavage site of the secretory target peptide is between position 25 and 26. Thus, the mature protein contains 118 amino acids with a total mass of 12 kDa, which is significantly more than the mass of a usual plant LTP. The size difference results from an N-terminal, proline-rich region that is absent from other LTPs. However, this N-terminal sequence is highly conserved between *Cg*LTP and its closest homologs in *A. thaliana* and *T. hassleriana*. To classify CgLTP as a family 1 or 2 LTP, which are 7 and 9 kDa proteins, respectively, we were thus not able to rely on the size of the mature protein. As an alternative approach, we tried to identify conserved sites that were identified for each LTP family in the *Cg*LTP sequence (Yeats 2008). Besides the highly conserved eight cysteine residues, we only found 3 of 13 family 1 residues and 4 of 12 family 2 residues conserved. Two of these conserved residues are shared between family 1 and 2. We were thus not able to classify CgLTP as a member of either LTP family 1 or 2. From our results we conclude that *Cg*LTP belongs to an unusual group of LTPs that commonly exists in *Brassicales* species.

Next, we verified the *Cg*LTP expression values that we extracted from RNA sequencing in *Cleome* species (Külahoglu et al., 2014) with semi-q RTPCR on cDNA of *C. gynandra* and *T. hassleriana*. PCR of *Cg*LTP yielded two to three times as much product as PCR of *Th*LTP despite similar size of the PCR products (94 and 100 bp, respectively). As an internal control, we amplified Actin7. The increased abundance of *Cg*LTP PCR product verifies our previous results in showing that expression of the analyzed LTP is highly up-regulated in the C_4 plant *C. gynandra*.

To assess, whether strong expression of *Cg*LTP has an effect on the lipid composition of *C. gynandra*, we compared the lipidome of *C. gynandra* and *T. hassleriana* via thin layer chromatography. We separated neutral lipids, phospholipids and glycolipids on separate chromatograms, identified lipid species with internal controls and interpolated remaining spots from previous analyses (Lepage, 1964; Vitello and Zanetta, 1978; Juguelin et al., 1986). The dominant lipid species we observed were monogalactosyl diacylglycerol (MGDGs) and digalactosyl diacylglycerols (DGDGs), which are highly abundant and important components of all photosynthetic membranes (Dörmann and Benning, 2002). On plates where neutral lipids were not separated, they also appeared as a major spot. However, upon separation no single abundant neutral lipid species, but several minor species were detected. The only apparent difference between the whole lipidome of both species is the lack of a fatty alcohol species in *C. gynandra*. The lack of this fatty alcohol is contrasted by larger accumulation of a second, smaller fatty alcohol species in the C_4 plant. We conclude that the lipidome between two closely

related species is quantitatively very similar with only slight differences, regardless of the mode of photosynthesis.

To functionally characterize *Cg*LTP, we analyzed its localization in leaf tissue with a combined approach of ß-glucuronidase assay and fluorescence labelling. GUS staining was visible exclusively in cells that are proximate to *A. thaliana* leaf veins and the yellow fluorescent Venus signal appeared in the cell periphery of transiently transformed *N. benthamiana* leaves. In combination with previous analyses of the secretory target peptide, these results indicate that *Cg*LTP is expressed in bundle sheath cells and secreted to the extracellular space.

Subsequently, we performed protein-lipid binding assays to determine substrates of *Cg*LTP. First we tested for binding to the most common hydrophobic components of lipids, namely fatty acids and sterols, as previously characterized LTPs were shown to bind to these hydrophobic metabolites (van Amerongen et al., 1989; Tsuboi et al., 1992). *Cg*LTP solely bound to cholesterol as indicated by a dark stain on the membrane, while it showed no binding affinity towards the other metabolites. To detect other putative binding partners of *Cg*LTP, we repeated the assay with an increased number of substrates. We included other abundant plant sterols, precursors of cuticular waxes and membrane lipids. To allow for the detection of very weak interactions, we furthermore excluded detergents from all washing steps in the protocol. The membrane thus shows a blurry background signal, but rarely any additional binding of other substrates than cholesterol. Staining of the cholesterol occured in a dose dependent manner, showing that binding strength can be quantified. Only one more lipid, namely campesterol showed a very faint, ring shaped signal, when 50 µg were applied to the membrane. Since this signal was weaker than the 10 µg signal from cholesterol, we conclude that campesterol binding occurs with less than 20% of cholesterol binding specificity. Campesterol is very similar to cholesterol. Both sterols differ only by single methyl group at C24 that is absent in cholesterol. These results led us to the conclusion that *Cg*LTP binds cholesterol in a specific manner with a minor side affinity towards the very similar campesterol. Our result shows that plant LTPs have the potential to specifically bind substrates like specific LTPs from other organisms (Sha and Luo, 1999). Interestingly, the fatty alcohol 1-Docosanol was not bound by *Cg*LTP. We therefore concluded that the differential regulation of *Cg*LTP is not directly causal for the changes in the fatty alcohol contents between *C. gynandra* and *T. hassleriana*.

To test whether sterol metabolism is affected by the strong expression of *Cg*LTP, we quantified free sterols via GC/MS. We also wanted to know, if putative effects are commonly seen in C_4 plants. Thus, we included a second pair of closely related C_3 and C_4 species, namely *F. bidentis* (C_4) and *F. robusta* (C_3), in the analysis. The sterol levels of all plants showed a similar trend: ß-sitosterol was the most abundant sterol, while the other sterols were less abundant. This observation is not unusual, as ß-sitosterol is the dominant sterol species in many plants (Grunwald, 1975; DOUGLAS and PALEG,

1981; Heupel et al., 1986; Wewer et al., 2011). The total content of free sterols thus depends mostly on the ß-sitosterol content. In the reverse conclusion, the increased ß-sitosterol content of *T. hassleriana* indicates increased total sterol content in the plant. This trend is, however, not seen in the C_3 plant *F. robusta* and we suggest that it is a species-specific effect rather than being related to the mode of photosynthesis. When we analyzed relative changes in sterol contents between C_3 and C_4 species, we found that *F. bidentis* and *C. gynandra* display largely opposite adaptations as compared to their closest related C_3 plants. We thus suggest that there is no common adaptation in sterol composition of C_4 plants. Previous studies on the lipidome of the C_4 species sorghum and maize is in line with our findings, since they share no common pattern of sterol compositions as compared to C_3 plants, too (DOUGLAS and PALEG, 1981; Heupel et al., 1986; Wewer et al., 2011). Accumulation of sterols was also shown to be involved in the adaptation to extreme environments (Kuiper, 1985; Nakamura et al., 2001). Thus, we interpreted the sterol content of *C. gynandra* in a species-specific manner and suggest that adaptation of sterol contents is not a prerequisite for C_4 photosynthesis. *C. gynandra* contains sparse amounts of cholesterol (1% of total free sterols) as compared to *T. hassleriana* (3%). This is in contrast to the high abundance of the cholesterol-binding *Cg*LTP. Multiple hypotheses can explain this antithetic abundance of cholesterol and a cholesterol binding protein. The interaction of CgLTP with cholesterol could be involved in cell-to-cell signaling as it was shown for an LTP involved in pathogen response (Maldonado et al., 2002). This hypothesis aligns well with the low content of cholesterol, which allows for high specificity of an intercellular signal. It could also contribute to the establishment of a local structure, like plasmodesmata. Remorin, a sterol binding protein, was shown to be enriched at the site of plasmodesmata (Raffaele et al., 2009). It was thus hypothesized that sterol rich domains are involved in plasmodesmata formation. Still, the ubiquitous expression of *Cg*LTP in the cell periphery contradicts a function in local processes. It is also less probable that *Cg*LTP integrates cholesterol in apoplastic structures like suberin or cutin, since cholesterol or other related metabolites were not identified as abundant components of these polymers in previous lipidomics studies (Pollard et al., 2008). However, a role of glycosylated sterols in cell wall initiation was proposed (Peng et al., 2002; Reiter, 2002). The lack of a phenotype in the UDP-Glucose:Sterol Glucosyltransferase contradict this hypothesis, though (DeBolt et al., 2009).

Since *Cg*LTP is expressed exclusively in bundle sheath cells, we tested whether a lack of its substrate cholesterol causes impairment in vein formation. Therefore, we applied Mevastatin, an inhibitor of cholesterol biosynthesis, to developing leaves of *C. gynandra* and measured vein density. We monitored a slight, insignificant decrease in vein density in treated leaves. We conclude that *Cg*LTP and its binding to cholesterol do not or only slightly affect vein development.

Summary

In this study, we characterize the function a highly abundant lipid transfer protein from the C_4 plant *C. gynandra*. CgLTP specifically binds cholesterol and is located to the apoplast of bundle sheath cells. Despite many indicators for a function of *Cg*LTP in C_4 bundle sheath development, a prediction of a biological function requires further work.

Outlook

To determine the function of *Cg*LTP *in vivo*, we will analyze the localization of the protein more precisely in GUS-stained cross sections of leaf veins and their surrounding tissue. For further functional characterization, analysis of *Cg*LTP knockdown plants of *C. gynandra* would provide an ideal model. Expression and purification of significant amounts of functional *Cg*LTP in *Escherichia coli* cells would further help biochemical characterization of the protein.

Supplemental information

Supplemental dataset 1: List of oligonucleotides

References

Arondel, V., and Kader, J.C. (1990). Lipid transfer in plants. Experientia **46,** 579-585.

Arondel, V., Vergnolle, C., Cantrel, C., and Kader, J.-C. (2000). Lipid transfer proteins are encoded by a small multigene family in Arabidopsis thaliana. Plant Science **157,** 1-12.

Bitman, J., and Wood, D. (1982). An improved copper reagent for quantitative densitometric thin-layer chromatography of lipids. Journal of Liquid Chromatography **5,** 1155-1162.

Boutrot, F., Chantret, N., and Gautier, M.-F. (2008). Genome-wide analysis of the rice and arabidopsis non-specific lipid transfer protein (nsLtp) gene families and identification of wheat nsLtp genes by EST data mining. BMC Genomics **9,** 86.

Cheng, S., van den Bergh, E., Zeng, P., Zhong, X., Xu, J., Liu, X., Hofberger, J., de Bruijn, S., Bhide, A.S., Kuelahoglu, C., Bian, C., Chen, J., Fan, G., Kaufmann, K., Hall, J.C., Becker, A., Brautigam, A., Weber, A.P., Shi, C., Zheng, Z., Li, W., Lv, M., Tao, Y., Wang, J., Zou, H., Quan, Z., Hibberd, J.M., Zhang, G., Zhu, X.G., Xu, X., and Schranz, M.E. (2013). The Tarenaya hassleriana genome provides insight into reproductive trait and genome evolution of crucifers. The Plant cell **25,** 2813-2830.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162,** 156-159.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana. The Plant Journal **16,** 735-743.

Coutos-Thevenot, P., Jouenne, T., Maes, O., Guerbette, F., Grosbois, M., Le Caer, J.P., Boulay, M., Deloire, A., Kader, J.C., and Guern, J. (1993). Four 9-kDa proteins excreted by somatic embryos of grapevine are isoforms of lipid-transfer proteins. European Journal of Biochemistry **217,** 885-889.

Deblaere, R., Bytebier, B., De Greve, H., Deboeck, F., Schell, J., Van Montagu, M., and Leemans, J. (1985). Efficient octopine Ti plasmid-derived vectors for Agrobacterium-mediated gene transfer to plants. Nucleic Acids Res. **13,** 4777-4788.

DeBolt, S., Scheible, W.-R., Schrick, K., Auer, M., Beisson, F., Bischoff, V., Bouvier-Navé, P., Carroll, A., Hematy, K., Li, Y., Milne, J., Nair, M., Schaller, H., Zemla, M., and Somerville, C. (2009). Mutations in

UDP-Glucose:Sterol Glucosyltransferase in Arabidopsis Cause Transparent Testa Phenotype and Suberization Defect in Seeds. Plant Physiol. **151,** 78-87.

Dörmann, P., and Benning, C. (2002). Galactolipids rule in seed plants. Trends Plant Sci. **7,** 112-118.

DOUGLAS, T.J., and PALEG, L.G. (1981). Lipid Composition of Zea mays Seedlings and Water Stress-Induced Changes. Journal of Experimental Botany **32,** 499-508.

Douliez, J.P., Michon, T., Elmorjani, K., and Marion, D. (2000). Mini Review: Structure, Biological and Technological Functions of Lipid Transfer Proteins and Indolines, the Major Lipid Binding Proteins from Cereal Kernels. Journal of Cereal Science **32,** 1-20.

Dowler, S., Kular, G., and Alessi, D.R. (2002). Protein Lipid Overlay Assay.

Griebel, T., and Zeier, J. (2010). A role for beta-sitosterol to stigmasterol conversion in plant-pathogen interactions. The Plant journal : for cell and molecular biology **63,** 254-268.

Grunwald, C. (1975). Phytosterols in tobacco leaves at various stages of physiological maturity. Phytochemistry **14,** 79-82.

Heupel, R., Sauvaire, Y., Le, P., Parish, E., and Nes, W.D. (1986). Sterol composition and biosynthesis in sorghum: Importance to developmental regulation. Lipids **21,** 69-75.

Inda, L., Torrecilla, P., Catalán, P., and Ruiz-Zapata, T. (2008). Phylogeny of Cleome L. and its close relatives Podandrogyne Ducke and Polanisia Raf. (Cleomoideae, Cleomaceae) based on analysis of nuclear ITS sequences and morphology. Plant Syst Evol **274,** 111-126.

Juguelin, H., Heape, A., Boiron, F., and Cassagne, C. (1986). A quantitative developmental study of neutral lipids during myelinogenesis in the peripheral nervous system of normal and trembler mice. Brain research **390,** 249-252.

Kader, J.-C. (1996). LIPID-TRANSFER PROTEINS IN PLANTS. Annu. Rev. Plant Physiol. Plant Molec. Biol. **47,** 627-654.

Kader, J.-C., Julienne, M., and Vergnolle, C. (1984). Purification and characterization of a spinach-leaf protein capable of transferring phospholipids from liposomes to mitochondria or chloroplasts. European Journal of Biochemistry **139,** 411-416.

Kader, J.C. (1975). Proteins and the intracellular exchange of lipids. I. Stimulation of phospholipid exchange between mitochondria and microsomal fractions by proteins isolated from potato tuber. Biochimica et biophysica acta **380,** 31-44.

Krebs, M., Held, K., Binder, A., Hashimoto, K., Den Herder, G., Parniske, M., Kudla, J., and Schumacher, K. (2012). FRET-based genetically encoded sensors allow high-resolution live cell imaging of Ca(2)(+) dynamics. The Plant journal : for cell and molecular biology **69,** 181-192.

Kuiper, P.J.C. (1985). Environmental changes and lipid metabolism of higher plants. Physiol. Plant. **64,** 118-122.

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

Lascombe, M.B., Bakan, B., Buhot, N., Marion, D., Blein, J.P., Larue, V., Lamb, C., and Prange, T. (2008). The structure of "defective in induced resistance" protein of Arabidopsis thaliana, DIR1, reveals a new type of lipid transfer protein. Protein science : a publication of the Protein Society **17,** 1522-1530.

Lepage, M. (1964). Isolation and characterization of an esterified form of steryl glucoside. Journal of Lipid Research **5,** 587-592.

Maldonado, A.M., Doerner, P., Dixon, R.A., Lamb, C.J., and Cameron, R.K. (2002). A putative lipid transfer protein involved in systemic resistance signalling in Arabidopsis. Nature **419,** 399-403.

Mishina, T.E., and Zeier, J. (2006). The Arabidopsis Flavin-Dependent Monooxygenase FMO1 Is an Essential Component of Biologically Induced Systemic Acquired Resistance. Plant Physiol. **141,** 1666-1675.

Nagaya, S., Kawamura, K., Shinmyo, A., and Kato, K. (2010). The HSP terminator of Arabidopsis thaliana increases gene expression in plant cells. Plant & cell physiology **51,** 328-332.

Nakamura, A., Shimada, H., Masuda, T., Ohta, H., and Takamiya, K.-i. (2001). Two distinct isopentenyl diphosphate isomerases in cytosol and plastid are differentially induced by environmental stresses in tobacco1. FEBS Lett. **506,** 61-64.

Norris, S.R., Meyer, S.E., and Callis, J. (1993). The intron of Arabidopsis thaliana polyubiquitin genes is conserved in location and is a quantitative determinant of chimeric gene expression. Plant Mol Biol **21,** 895-906.

Ohlrogge, J., and Browse, J. (1995). Lipid biosynthesis. The Plant cell **7,** 957-970.

Peng, L., Kawagoe, Y., Hogan, P., and Delmer, D. (2002). Sitosterol-β-glucoside as Primer for Cellulose Synthesis in Plants. Science **295,** 147-150.

Pollard, M., Beisson, F., Li, Y., and Ohlrogge, J.B. (2008). Building lipid barriers: biosynthesis of cutin and suberin. Trends Plant Sci. **13,** 236-246.

Raffaele, S., Bayer, E., Lafarge, D., Cluzet, S., German Retana, S., Boubekeur, T., Leborgne-Castel, N., Carde, J.-P., Lherminier, J., Noirot, E., Satiat-Jeunemaître, B., Laroche-Traineau, J., Moreau, P., Ott, T., Maule, A.J., Reymond, P., Simon-Plas, F., Farmer, E.E., Bessoule, J.-J., and Mongrand, S. (2009). Remorin, a Solanaceae Protein Resident in Membrane Rafts and Plasmodesmata, Impairs Potato virus X Movement. The Plant Cell Online **21,** 1541-1555.

Reiter, W.-D. (2002). Biosynthesis and properties of the plant cell wall. Current Opinion in Plant Biology **5,** 536-542.

Romeis, T., Ludwig, A.A., Martin, R., and Jones, J.D.G. (2001). Calcium-dependent protein kinases play an essential role in a plant defence response. The EMBO Journal **20,** 5556-5567.

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

Sha, B., and Luo, M. (1999). PI transfer protein: the specific recognition of phospholipids and its functions. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids **1441,** 268-277.

Sheen, J. (1991). Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes. The Plant Cell Online **3,** 225-245.

Sterk, P., Booij, H., Schellekens, G.A., Van Kammen, A., and De Vries, S.C. (1991). Cell-specific expression of the carrot EP2 lipid transfer protein gene. The Plant Cell Online **3,** 907-921.

Tchang, F., This, P., Stiefel, V., Arondel, V., Morch, M.D., Pages, M., Puigdomenech, P., Grellet, F., Delseny, M., and Bouillon, P. (1988). Phospholipid transfer protein: full-length cDNA and amino acid sequence in maize. Amino acid sequence homologies between plant phospholipid transfer proteins. Journal of Biological Chemistry **263,** 16849-16855.

Thoma, S., Kaneko, Y., and Somerville, C. (1993). A non-specific lipid transfer protein from Arabidopsis is a cell wall protein. The Plant Journal **3,** 427-436.

Tsuboi, S., Osafune, T., Tsugeki, R., Nishimura, M., and Yamada, M. (1992). Nonspecific Lipid Transfer Protein in Castor Bean Cotyledon Cells: Subcellular Localization and a Possible Role in Lipid Metabolism. Journal of Biochemistry **111,** 500-508.

van Amerongen, A., Demel, R.A., Westerman, J., and Wirtz, K.W.A. (1989). Transfer of cholesterol and oxysterol derivatives by the nonspecific lipid transfer protein (sterol carrier protein 2): A study on its mode of action. Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism **1004,** 36-43.

Veerkamp, J.H., Peeters, R.A., and Maatman, R.G.H.J. (1991). Structural and functional features of different types of cytoplasmic fatty acid-binding proteins. Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism **1081,** 1-24.

Vitello, F., and Zanetta, J.P. (1978). Thin-layer chromatography of phospholipids. Journal of Chromatography A **166,** 637-640.

Vitha, S. (2007). Histochemical localization of β-glucuronidase (GUS) reporter activity in plant tissues. Microscopy and Imaging Center.

Walter, M., Chaban, C., Schutze, K., Batistic, O., Weckermann, K., Nake, C., Blazevic, D., Grefen, C., Schumacher, K., Oecking, C., Harter, K., and Kudla, J. (2004). Visualization of protein interactions in

living plant cells using bimolecular fluorescence complementation. The Plant journal : for cell and molecular biology **40,** 428-438.

Wewer, V., Dombrink, I., vom Dorp, K., and Dörmann, P. (2011). Quantification of sterol lipids in plants by quadrupole time-of-flight mass spectrometry. Journal of Lipid Research **52,** 1039-1054.

Wirtz, K.W.A. (1991). Phospholipid Transfer Proteins. Annu. Rev. Biochem. **60,** 73-99.

Yamada, M. (1992). Lipid Transfer Proteins in Plants and Microorganisms. Plant Cell Physiol. **33,** 1-6.

Yeats, T.H., and Rose, J.K.C. (2008). The biochemistry and biology of extracellular plant lipid-transfer proteins (LTPs). Protein Science **17,** 191-198.

Manuscript 5

Plasticity of C_4 photosynthesis in the amphibious sedge *Eleocharis retroflexa*

Canan Külahoglu[a], Simon Schliesky[a], Manuel Sommer[a], Alisandra K. Denton[a], Andreas Hussner[b], C. Robin Buell[c], Andrea Bräutigam[a] and Andreas P. M. Weber[a1]

[a]Institute of Plant Biochemistry, Cluster of Excellence on Plant Sciences, Heinrich-Heine-University, 40225 Düsseldorf, Germany

[b]Institute of Plant Biochemistry- Photosynthesis and Stress Physiology of Plants, Cluster of Excellence on Plant Sciences, Heinrich- Heine-University, 40225 Düsseldorf, Germany

[c]Department of Plant Biology, Michigan State University, 48824 East Lansing, MI , USA

[1]Corresponding author; e-mail andreas.weber@uni-duesseldorf.de.

Abstract

C_4 photosynthesis is a complex adaptive trait, facilitating the adaption to hot and arid environments. It has been hypothesized that evolution of the C_4 trait comes at the cost of reduced phenotypical plasticity, owing to the complex anatomical and biochemical specialization required for operating the C_4 carbon concentrating mechanism. However, C_4 photosynthetic terrestrial wetland species of the genus *Eleocharis* display a remarkable phenotypic plasticity in their mode of photosynthetic carbon assimilation. In particular *Eleocharis retroflexa*, which is classified as the most C_4-like species amongst the known *Eleocharis* C_4 performing sedges, is able to thrive under submerged conditions by reconfiguration of its culm anatomy and potentially its photosynthetic mode. The underlying molecular mechanisms permitting adaptation to environmental change through metabolic plasticity are however unknown to date. To begin to unravel the physiological and transcriptional programs that enable *E. retroflexa* to thrive during submergence and on soil, we employed deep RNA-sequencing of aquatically and terrestrially grown culms and contextualized these molecular data with physiological parameters and enzyme activity measurements. We found that *E. retroflexa* undergoes structural and metabolic rewiring during submergence by adapting its culms fully to the new habitat and adjusting its carbon metabolism. While the aquatic *E. retroflexa* culm transcriptome reflects characteristics described for flooding tolerant plants, the carbon metabolism displays a typical C_3-C_4 intermediate signature, featuring high abundance of photorespiratory transcripts. At the same time the C_4 cycle is maintained. Owing to its metabolic plasticity, *E. retroflexa* represents an interesting model to unravel the molecular mechanisms of adaptation to changing environments by phenotypic plasticity.

Introduction

Phenotypic plasticity describes the ability of organisms to accommodate and react to variable environmental conditions by changing their characteristics for better acclimatization (Pigliucci 2001; Sage and McKown 2006).

The CO_2 concentrating mechanism of C_4 photosynthesis is considered as a specialized adaptation derived from C_3 ancestors. It is a complex trait, which is employed for carbon gain in hot, often arid and high light environments to circumvent high photorespiration rates. The high degree of anatomical and biochemical specialization of C_4 photosynthesis performing plants is thought to reduce their potential for phenotypic plasticity and photosynthetic acclimation to variable environments, as compared to C_3 plants (reviewed by Sage and McKown 2006).

C_4 photosynthesis requires a distinct anatomical and biochemical infrastructure for optimal functionality (Hatch 1987). Generally spoken, the C_4 pathway acts as a carbon concentrating mechanism that works on top of the C_3 photosynthetic carbon assimilation by increasing the local CO_2 concentration next to the ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO; Bowes et al. 1971; Furbank and Hatch 1987). Typically, with few exceptions, C_4 leaves have two types of photosynthetic cells, with carbon in the form of HCO_3^- initially fixed by the phospho*enol*pyruvate carboxylase (PEPC) in the outer mesophyll cells (MCs) and then shuttled in the form of a C_4 carbon compound into the inner bundle sheath cells (BSCs; Hatch and Slack 1970). In the BSCs the C_4 carbon compound is decarboxylated, releasing CO_2 at the site of the RuBisCO by either the NAD-dependent malic enzyme (NAD-ME), the NADP-dependent malic enzyme (NADP-ME), or the phospho*enol*pyruvate carboxykinase (PEPCK), followed by assimilation into carbohydrates by the Calvin-Benson-Bassham cycle (CBBC; Hatch 1987; Hatch and Slack 1970). The remaining C_3 molecule is transported back to the MCs. The BSCs are situated outside the vascular bundle, encompassing it like a wreath, which is termed “Kranz”-anatomy (Haberlandt 1904).

The above-described C_4-specific coordinated modifications of both metabolism and anatomy may have reduced the ability of C_4 performing plants to acclimate their photosynthetic apparatus to altering environments (reviewed by Sage and McKown, 2006).

The sedge family (*Cyperaceae*) contains more than 20% of the currently known C_4 plant species (Besnard et al. 2009; Sage 2004). Among those terrestrial wetland species, members of leafless *Eleocharis* (Cyperaceae) genus display a remarkable degree of acclimation to varying habitats. These species can grow underwater as well as in air (Ueno 2001; Ueno et al. 1989; Ueno et al. 1988). Among the amphibious *Eleocharis* species (e.g. *E. retroflexa*, *E. vivipara*, *E. baldwinii*) the photosynthetic modes can be highly variable between aquatic and terrestrial habitat (Ueno 2004; Ueno and Wakayama 2004). While the culms of the terrestrial form of the three species show a C_4 photosynthesis signature of the NAD-ME subtype and Kranz anatomy, in the aquatic environment the culms of *E. retroflexa*, *E. vivipara*, *E. baldwinii* appear more C_4-like, C_3-C_4 intermediate, or C_3, respectively (Ueno 2004). Upon flooding, *E. retroflexa* culms undergo acclimatization of the terrestrial culms within days, while new aquatic adapted culms grow (Ueno and Wakayama 2004). However, the aquatic culms, when exposed to air, die away due to rapid drying, while new C_4 terrestrial accustomed culms grow out, as reported for other amphibious *Eleocharis* species (Ueno 2001; Ueno et al. 1988). Different strategies are known, which submerged terrestrial wetland species are employing to overcome the new challenges of the aquatic environment. The challenges of the aquatic environment are, for example, physical restrictions on light availability, gas exchange, and nutrient availability (Krause-Jensen and Sand-Jensen 1998; Pedersen et al. 2013). In particular, the gas diffusion rates in

water are 10^4-fold slower than in air. The limitation of CO_2 and O_2 gas flow severely affects photosynthesis and likely promotes increased rates of photorespiration. Also light intensity is subdued in turbid flooding water, which decreases photosynthesis efficiency further (Vervuren et al. 2003). The resulting imbalance between carbohydrate assimilation and consumption has lethal consequences for most flooding non-adapted terrestrial plants (Colmer and Voesenek 2009). As an adaptation to flooding, some species grow out of water by shoot elongation to reestablish aerial photosynthesis (Setter and Laureles 1996), whereas others develop specialized “aquatic leaves” specifically accustomed to the wet habitat (Bailey-Serres and Voesenek 2008). Aquatic adapted leaves display reduced gas diffusion resistance as a consequence of reduced cuticle thickness, chloroplast reorientation close towards the epidermis, and reduced leaf thickness (Frost-Christensen et al. 2003; Mommer et al. 2006; Mommer et al. 2005b; Sand-Jensen and Frost-Christensen 1999).

Previous studies indicated that *E. retroflexa* might have the capability to change its photosynthetic mode from C_4 to C_4-like, depending on the habitat (Ueno and Wakayama 2004). In this study we address, which physiological and transcriptional programs enable *E. retroflexa* to thrive during submergence and on land, and how the change of environment affects the photosynthetic modes (C_4-like and C_4 photosynthesis).

Material and Methods

Plant material and cultivation

Plants were purchased from an online aquarist-shop (http://www.wasserflora.de; B030PP). *E. retroflexa* plants were cultivated in terrestrial and aquatic culture for transcriptome profiling by Illumina Sequencing between January and March 2012. Terrestrial *E. retroflexa* cultures were grown on turf soil in boxes that were partially flooded with tap water under greenhouse conditions (21°C, 12:12h of light/darkness). For the drought stress experiment *E. retroflexa* seedlings were transferred to soil and grown for 14 days under the experimental conditions (group 1: control, every day 250 ml water; group2: every two days 250ml water; group3: every four days 150ml water; group4: no water for 14 days). The aquatic culture was set up by transferring viviparous plantlets to aquariums covered with turf soil and a 3 cm upper layer of gravel. Aquariums were filled with tap water and algal growth was suppressed by co-cultivation of shrimp. Temperature in aquariums was constant at approx. 25°C and fresh-air was constantly supplemented by an aquarium pump. Culms of aquatic and terrestrial culture were harvested after four weeks of growth. Up to 20 individual plants were pooled for each biological replicate.

Internal transcribed spacer sequence analysis and phylogeny

Internal transcribed spacer (ITS) sequence of *E. retroflexa* were subcloned with ITS1 and ITS4 primer (according to Inda et al., 2008) and sequenced. Plant identity was confirmed by comparison to public database Genbank. Sequenced *E. retroflexa* ITS was aligned with other known ITS from species of the *Eleocharis* clade by Clustal W (Larkin et al. 2007).

RNA extraction, library construction and sequencing

Plant material was extracted with 65°C pre-heated CTAB buffer working solution (1ml buffer per 100mg ground tissue): 50 % (v/v) CTAB buffer stock solution, 2 % (v/v) BME, and 50 % (v/v) acidic phenol. Ground tissue was incubated for 20min at 65°C, followed by two consecutive protein extractions by adding equal volume chloroform-isoamyl alcohol (24:1) to extract and 20 min centrifugation at 10,000xg at 10°C. The aqueous supernatant was transferred to fresh reaction tube and 0.5 volumes of 96% (v/v) ethanol were added. This mixture was loaded onto RNA binding silica columns (Plant RNeasy extraction kit; Qiagen, Hilden, Germany) and further processed as recommend by the manufacturer. RNA was treated twice with RNAse-free DNAse, first on-column and after elution a second time in solution (New England Biolabs, MA, USA). RNA integrity, sequencing library quality, and fragment size were checked on a 2100 Bioanalyzer (Agilent, CA, USA). Libraries were prepared using the TruSeq RNA Sample Prep Kit v2 (Illumina, San Diego, CA) and library quantification was performed with a Qubit 2.0 instrument (Invitrogen, Germany). Single end sequenced samples were multiplexed (6 libraries per lane with approximately 20 million reads per library). All libraries were sequenced on the HISEQ2000 Illumina platform (San Diego, CA). Libraries were sequenced in the single-end mode with read lengths ranging from 80-100 nucleotides.

Transcriptome assembly and annotation

Reads were checked for quality with FASTQC (http://bioinformatics.babraham.ac.uk/projects/fastqc/) and subsequently cleaned and filtered for quality scores greater than 20 and read length greater than 50 nucleotides using the FASTX toolkit (Blankenberg et al. 2010) http://hannonlab.cshl.edu/fastx_toolkit). Read files and assembly are deposited on data medium enclosed to this thesis..

Trimmed reads were split into subgroups and were assembled by CAP3 (Huang and Madan 1999). The resulting contigs were merged, split into subgroups again and assembled by CAP3. With this second assembly step, contigs and singlets were merged and assembled in CAP3. The resulting *E. retroflexa* filtered unigene database (contigs>200bases length) was annotated using BLASTX searches (cut-off $1e^{-10}$) against the *S. italica* primary transcript database V2.1 (Bennetzen et al. 2012) and Uniref100 (Bairoch et al. 2005). The best blast hit per read was filtered by the highest bitscore. Multiple matching contigs to one *S. italica* identifier were filtered out with customized Perl script.

Unigene database was filtered for contigs that either match an *S. italica* identifier or a plant identifier in Uniref100. This resulted in a unigene database of 27,021 contigs and reduced possible contamination through non-plant contigs.

The final unigene database was uploaded to the KAAS server (http://www.genome.jp/tools/kaas/) to test the representation of KEGG annotated pathways (Moriya et al. 2007). Resulting maps were manually curated for pathways present in plants by comparison to model species *A. thaliana* and analyzed for coverage.

Gene expression profiling

Expression abundances were determined by mapping the single-end read libraries (each replicate for each condition) independently against *S. italica* primary transcript coding sequences V2.1 (Bennetzen et al. 2012) using BLAT V35 (Kent 2002) in dnax mode (nucleotide sequence of query and reference are translated in six frames to protein) and counting the best mapping hit based on e-value for each read uniquely. Default BLAT parameters were used for mapping. Expression was normalized to reads per million mappable reads (RPM). A threshold of 20 RPM per transcript in at least one condition present in at least one replicate was chosen to discriminate against background transcription. Differential expressed transcripts were determined via EdgeR (Robinson et al. 2010) in R (R Development Core Team, 2009). A significance threshold of 0.05 was applied after *P*-value was adjusted for the False Discovery Rate (FDR) via Bonferroni-Holms correction (Holm 1979).

Cross species mapping sensitivity assessment

Each *E. retroflexa* read library was mapped to the unigene database and to the *S. italica* reference by Blat as described above (*see* **Gene expression profiling**). Raw read count files were sorted descending by number of aligned reads per identifier mapped reads and the amount of reads relative to all mapped reads per sample was summed up in R and plotted on a log10 scale. For comparison with other species and cross species mapping against own genome mapping, we mapped (i) *T. hassleriana* mature leaf reads (Külahoglu et al. 2014) against the *Arabidopsis thaliana* TAIR10 representative gene models (Lamesch et al. 2012) and (ii) *T. hassleriana* mature leaf reads against its own gene models (Cheng et al. 2013).

Data analysis

Data analysis was performed with the R statistical package (R Development Core Team, 2009) and Multi Experiment Viewer 4 (MEV4; http://www.tm4.org/mev/; Saeed et al. 2006; Saeed et al. 2003) unless stated otherwise. Before Principal component analyses (PCA) with median centering, the sample averages were z-score normalized. Hierarchical clustering of samples was performed with

MEV4 by normalizing them to z-scores and clustering with average linkage in Euclidean Distance. Sample enrichment was tested for tissue 'signature genes' with expression over 1,000 RPM in each tissue using *S. italica* V2.1 Mapman categories (from http://mapman.gabipd.org). Significantly differentially expressed transcripts (FDR<0.05) were tested for enrichment by Fisher's Exact Test and *p*-values were adjusted to FDR via Benjamini-Yekutieli correction (Yekutieli and Benjamini 1999). Mapman fold-change heatmaps were generated using the latest MAPMAN tool V3.6 with the *Setaria italica* V2.1 as reference (Thimm et al. 2004; Usadel et al. 2005). The Wilcoxon rank test was used for testing significance of fold-changes between the averages of aquatic and terrestrial transcriptomes for specific data subsets, with Benjamini-Yekutieli FDR correction of *P*-values (Usadel et al. 2005).

Culm anatomy analysis

Fresh culms of *E. retroflexa* grown submerged and on soil were cut transversally and imaged with the fluorescence microscope Axio Imager M2M (Zeiss, Germany) with light and fluorescence using an UV filter. Images were processed with ZEN10 software (Zeiss, Germany).

Quantitative real-time PCR

Quantitative real time PCR (qRT-PCR) was performed with three biological and three technical replicates per sample using the relative quantification technique by normalizing the gene of interest to a house-keeping gene (UBQ10). SYBR-green (MESA GREEN qPCR MasterMix Plus; Eurogentec) and gene specific primers (Supplemental Table 1) were employed as described by Schmittgen and Livak (2008). Mean normalized expression (MNE) was calculated via the ΔΔCT method after Pfaffl (2001).

Enzyme activity and chlorophyll measurements

For the enzyme activity assays under water stress, *E. retroflexa* terrestrial culms were grown with decreasing amounts of water in four biological replicates for two weeks (*see* **Plant material and cultivation**). Enzymatic activities of PEPC, PEPCK, AlaAT, AspAT and NAD-ME were determined as summarized by Ashton et al. (1990) in three biological replicates with three technical replicates per sample. Chlorophyll measurements were performed according to Porra et al (1989) with three biological replicates and three technical replicates per sample of aquatically and terrestrially grown culms.

Carbon isotope discrimination

For ^{13}C isotope discrimination leaf powder was freeze-dried and analyzed using the isotope ratio mass spectrometer IsoPrime 100 (ISOTOPE cube; Elementar Analysensysteme). Results were expressed

as relative values compared to the international standard (Vienna Pee Dee Belemnite) Element Analysis and calibration for δ^{13C} measurements followed the two-point method described by Coplen et al. (2006).

Results

Sequencing and assembly of aquatic and terrestrial *E. retroflexa* libraries provides a unigene database covering most of the plant relevant pathways present in KEGG

To provide a reference transcriptome for further studies three biological replicates of terrestrial culms and two biological replicates of aquatic culms yielding between 29 and 21 million high quality reads, were obtained by Illumina RNA-sequencing (Table 1; Dataset 1). Due to the absence of a reference genome, reads were *de novo* assembled with CAP3 (Huang and Madan 1999), producing a contig database of 43,817 unigenes with an N50 of 984 bases (unigene length>200 bases length; Supplemental Table 2; Supplemental Figure 1A). Fifty-eight percent (25,386) of the *E. retroflexa* contigs were annotated by mapping them to the evolutionary closest available C_4 grass genome of *Setaria italica* (Bennetzen et al. 2012), matching to 37% (13,204) of the known *S. italica* genes (Table 1). To estimate the quality of the contigs and contamination due to co-cultivation in the aquaria, the contigs were annotated against the Uniref100 protein database (Bairoch et al. 2005; Supplemental Table 2). Out of 43,817 contigs, 29,512 (67%) found a best match in the Uniref100 database (Figure 1A). From these 21,832 (50%) were annotated to a *Viridiplantae* identifier (ID) by best hit (Figure 1A). Around 7,112 (16%) contigs fell into the category of non-plant annotated IDs matching to fungi, bacteria or insects (Figure 1A). In the subsequent analyses, the unigene database was limited to 27,021 contigs matching either *S. italica* identifier, plant identifier of Uniref100 or both with an N50 of 1199 bases (Dataset 2; Supplemental Figure 1B; Supplemental Table 2).

To assess whether core plant metabolism was well represented by the filtered *E. retroflexa* unigene database (27,021 contigs), it was benchmarked against plant pathways from KEGG (Moriya et al. 2007; Supplemental Dataset 1). The contig database covered all genes involved in light and dark reactions of photosynthesis, as well as starch and sucrose metabolism, tricarboxylic acid cycle (TCA cycle), glycolysis, galactose metabolism, pyruvate metabolism, amino sugar and nucleotide sugar and nucleotide metabolism. Other pathways of carbohydrate metabolism, such as the pentose phosphate pathway, fructose and mannose metabolism, and glyoxylate and dicarboxylate metabolism lacked full coverage by few genes (Supplemental Dataset 1). In general, the metabolism of lipids, amino acids and nucleotides were fully represented.

Secondary metabolism involving synthesis of lignin precursors derived from phenylpropanoid, carotenoid, flavonoid, and porphyrin and chlorophyll biosynthesis were completely covered. Terpenoid and anthocyanin synthesis were incomplete.

In summary, the presented *E. retroflexa* unigene database exhibited good coverage of all core plant metabolic pathways (Supplemental Dataset 1) as well as central cellular processes (DNA repair, DNA transcription, translation and protein; Supplemental Dataset 2), including regulatory networks and plant hormone signaling (Supplemental Dataset 3).

Mapping of *E. retroflexa* reads to *S. italica* improves transcriptome representation relative to mapping the reads to *E. retroflexa* contigs.

To quantify gene expression, the reads were mapped to a reference sequence. Two options are available for this –mapping of reads (i) to the *de novo* assembled contigs or (ii) to the genome of a related species. For *E. retroflexa* transcript quantification using both approaches, mapping to the contigs or a cross-species reference database, were compared and evaluated. At least 70% of the reads mapped to the unigene database (Table 1), however, annotation of the unigene database with *S. italica* revealed known issues of *de novo* assemblies (Schliesky et al. 2012). Around 58% of the annotated *E. retroflexa* unigenes were matched to a *S. italica* identifier, which was assigned as best hit to more than one unigene (Figure 1B; Supplemental Dataset 4). In a more extreme case, 100 contigs matched one *S. italica* identifier (Si020831m**)** encoding a protein of unknown function. C_4 cycle genes, such as the alanine aminotransferase (AlaAT) and the triose-phosphate transporter (TPT) were absent in the unigene database. For comparison *E. retroflexa* reads were aligned to the *S. italica* gene models using the Blat algorithm. In total, 21,679 *S. italica* identifiers were matched by cross species mapping of *E. retroflexa* reads, with mapping efficiencies between 28-36% for all mapped samples (Table 1). To compare both mapping approaches the fraction of reads mapping to higher and lower expressed sequences was visualized for all samples. Mapping the reads to *S. italica* gene models delivered higher similarity between the individually mapped biological replicates than mapping to the *E. retroflexa* unigenes (Supplemental Figure 2A). When mapping *E. retroflexa* reads to its own unigene database, on average 20% of all reads matched to four of the most highly expressed unigenes (Supplemental Figure 2A), resulting in the high starting point of the curve displayed in Supplementary Figure 2A. These contigs were annotated as plant specific 16S, 18S and 26S rRNA subunits. On the basis of these results and previous experience (Bräutigam et al. 2010; Gowik et al. 2011), we opted to conduct transcript quantification by cross-species mapping of *E. retroflexa* to *S. italica*.

The *E. retroflexa* transcriptomes reflect minor changes between different habitats and display

unexpected variability between replicates

For analyzing the degree of variation and gene expression dynamics of the *E. retroflexa* transcriptomes all samples were hierarchically clustered (Figure 2A) and reduced to their main variances by principle component analysis (PCA; Figure 2B). Biological replicates of the aquatic and terrestrial *E. retroflexa* culm transcriptomes clustered together and were separated by habitat (Figure 2A). The PCA reflected the hierarchical clustering, with the first component separating the samples by habitat, accounting for 36% of total sample variation and the second component with 21% describing the biological variation within the biological replicates (Figure 2B). In contrast Pearson's correlation between the biological replicates (average mean r = 0.87, Supplemental Table 3) was similar to the Pearson's correlation between averages of the transcriptomes of the two conditions (average mean r=0.86; Supplemental Table 3). These results indicated that one third of the gene expression was changed by the culm's habitat. Despite this environmental influence all samples were qualitatively similar with regard to the Pearson's correlation (Supplemental Table 3), but showed a constant factor of variability between all replicates. To test whether the observed variability between replicated samples was random or resulted from unintended variation of experimental conditions, we compared the variation between the genes called as differentially expressed and the remaining gene set. In total, 8% of the whole transcriptome (1,356 genes) was differentially regulated between the aquatic (630 up-regulated; BH corrected P-value<0.05) and terrestrial (726 up-regulated; BH corrected P-value<0.05) culms (Supplemental Figure 3). If the variation among the samples was random, it would be randomly distributed between the significantly changed transcripts and the remaining transcripts. We compared the variability among replicates of all 708 significantly changed transcripts. There was a significant enrichment between genes with two fold variation between replicates and genes that were differentially expressed genes (Fisher's exact test P-value<0.001; Figure 2C). Thus, the genes that were related to habitat acclimatization were the ones showing greater variation between replicates than other genes.

Functional changes in the aquatic and terrestrial culm transcriptomes mirror the acclimatization of *E. retroflexa* to the respective habitats

The terrestrial *E. retroflexa* culms express more transcripts related to structure

Aquatic *E. retroflexa* culms grew fast under water, but never lifted themselves beyond the water surface. This is similar to what has been described for the growth habitus of aquatic *E. vivipara* plants (Supplemental Figure 4; Ueno 2001). Thus, to trace down the differences between aquatic and terrestrial culm structure the respective transcriptomes were analyzed for pathway enrichment of cell wall related categories (Figure 3A). All mentioned changes were statistically tested for significance.

The aquatic culms invested much less in transcription of genes related to phenolic compounds, phenylpropanoid and lignin biosynthesis (Figure 3A). Transcripts of these categories were up-regulated and enriched in the terrestrial culms (Fisher's Exact BY corrected *P*-value 4.2E-4; Figure 3A; Supplemental Figure 5; Supplemental Dataset 6). The higher transcript abundance of structure-related genes in terrestrial culms was reflected by higher fold-changes of the category "cellulose synthesis for cell wall enforcement" (*SEC61 BETA, CELLULOSE SYNTHASE LIKE D4*; Wilcoxon rank test BY corrected *P*-value 0.014; Supplemental Dataset 7) as well as enrichment of fold changed transcript levels regarding cell wall modification, e.g. several classes of pectin esterases (*PECTIN ESTERASE 11, PECTIN METHYLESTERASE, QUARTET*) known to cause cell wall stiffening (Wilcoxon rank test BY corrected *P*-value 1.69E-4; Supplemental Dataset 7; Micheli 2001). There were at least two-fold-changes in expression levels of transcripts involved in cell wall loosening and expansion (Wilcoxon rank test BY corrected *P*-value 1.79E-10), such as glycosyl hydrolases, beta-xylosidases (*BETA-XYLOSIDASE 2* and *3)* endotransglucolases (*XYLOGLUCAN ENDOTRASNGLYCOSE 3*, *4* and *8)* and polygalacturonases (*POLYGALACTURONASE1* and *3* and *QUARTET2*), in the terrestrial culms (Figure 3A; Supplemental Dataset 7).

Within the Mapman category of cell wall modification (Wilcoxon rank test BY corrected *P*-value 0.0016; Supplemental Dataset 7) we could monitor up to five-fold-changed expression levels of transcripts annotated as expansins (e.g. *EXPANSIN7, EXPANSIN8, EXPANSIN11* and *EXPANSIN16*; Figure 3A). Further the terrestrial culms exhibited up-regulation of transcripts belonging to wax synthesis needed for cuticle development (Wilcoxon rank test BY corrected *P*-value 0.03; Figure 3A; Supplemental Dataset 7). These differences in structure related transcripts were reflected by a 2.7-fold difference in dry weight to fresh weight ratio between the aquatic and terrestrial culms (Figure 3B).

Comparing aquatically and terrestrially grown *E. retroflexa* culms revealed evident changes in culm anatomy (Figure 4A). The terrestrial culms showed higher auto-fluorescence of lignified tissue (xylem) under UV light than the aquatic culms (Figure 4B). Also the terrestrial BSCs had weak lignification (Figure 4B). The terrestrial epidermis was regularly interspersed with stomata whose cell walls displayed auto-fluorescence, as well as the auto-fluorescence of the cuticle waxes. In the aquatic culms no stomata could be detected. Thus, the anatomical and structural changes within culm anatomy were traceable in the transcriptome.

The photosynthesis apparatus is enhanced in aquatically grown culms

One of the challenges for photosynthesis under water is low light availability and the quality of the available light spectrum (Kirk 1994; Pedersen et al. 2013). We hence assessed the transcriptomes for consequences of the submerged lifestyle on photosynthesis and light capture. Most of the Mapman

annotated transcripts for light reactions including photosystem I and II polypeptide subunits, and light harvesting complexes, as well as the cytochrome b6f/c were up-regulated in the aquatic form (Wilcoxon rank test BY corrected *P*-value 1.09E-13; Figure 4A; Supplemental Dataset 7). Analysis of cumulative gene expression showed that light reactions occupy 10% of transcriptional investment (Supplemental Figure 6). Synthesis of glycolipids in general was up-regulated (Fisher's exact test BY corrected *P*-values <0.001; Supplemental Dataset 6). Transcripts associated with the biosynthesis of thylakoid membrane lipids, such as transcripts of *DIGALACTOSYL DIACYLGLYCEROL DEFICIENT 1* and *2* (*DGD1* and *2*) and *SULFOQUINOVOSYLDIACYLGLYCEROL 1* and *2* (*SQD1* and *2*), were up-regulated in the aquatic culms (Figure 4A, Supplemental Dataset 7). Concordantly with the enrichment of light harvesting complexes in submersed culms tetrapyrrole biosynthesis was up-regulated (Wilcoxon rank test BY corrected *P*-value 1.52E-4; Figure 5A; Supplemental Dataset 7). Notably, in the aquatic culms transcripts associated with chlorophyll and carotenoid biosynthesis were twice as abundant as in the terrestrial culms (Figure 5A; Supplemental Dataset 7).

As indicated by the transcriptomes, photometric chlorophyll determination revealed that total chlorophyll content per dry weight was two-fold higher in the aquatic culms (*P*-value<0.05; Figure 5B). The transcriptome changes associated with light capture and chloroplasts were reflected in culm anatomy. In terrestrial culms the outer BSCs were enlarged and accumulated high numbers of chloroplasts as seen in most C_4 plants. The MCs appeared much smaller (Figure 4A). Culms grown under submerged conditions featured less enlarged BSCs, while MC size and chloroplast number increased, compared to terrestrial MC culm anatomy (Figure 4A).

The C_4 cycle signature is stronger in terrestrial culms, while aquatic culms display enhanced expression of the Calvin-Benson-Bassham cycle and photorespiration

In addition to the detected changes in culm structure and light capture, the central carbon metabolism was adapted to the aquatic and terrestrial habitats (Figure 5A). Ueno and colleagues previously analyzed the localization of C_4 cycle enzymes (PEPC, NAD-ME, PPDK and RuBisCO Large Subunit) and classified *E. retroflexa* as NAD-ME subtype C_4 plant (Ueno and Wakayama 2004). The transcriptomes of the terrestrial and aquatic culms now enabled a detailed analysis of *E. retroflexa*'s full C_4 cycle and carbon concentrating mechanism under different growth conditions. The C_4 cycle showed clear differences between the terrestrial and aquatic habitats regarding transcript levels of C_4 cycle genes typical for the NAD-ME/PEPCK subtype (Figure 6A). The mentioned C_4 cycle genes were expressed between 593-25,621 RPM in the terrestrial and between 234-9,972 RPM in the aquatic culms (Supplemental Table 4). In the aquatic culms, abundance of C_4 cycle genes was between 31 to 83% of the terrestrial expression (Figure 6A; Supplemental Table 4). The three-transporter system

*BILE ACID:SODIUM SYMPORTER FAMILY PROTEIN2/SODIUM:HYDROGEN ANTIPORTER/PHOSPHO*ENOL*PYRUVATE TRANSLOCATOR* (*BASS2/NHD/PPT*), importing substrates (phosphate and pyruvate) for pyruvate, phosphate dikinase (PPDK) activity and exporting phospho*enol*pyruvate (PEP) in the MC was highly abundant in the terrestrial culms (Figure 6A). The enzymes needed for providing the substrate for HCO_3^- fixation, *PPDK* (regenerating PEP from pyruvate), a cytosolic *CARBONIC ANHYDRASE* (*CA2*; converting CO_2 to HCO_3^-), and *PEPC* (converting HCO_3^- and PEP to oxaloacetate; OAA) were highly abundant in the terrestrial compared to the aquatic culms. The transcripts encoding the main decarboxylating enzymes of this C_4 cycle subtype, *NAD-ME* and also *PEPCK,* were relatively higher in the terrestrial culms (Figure 6A). Also the transcript levels of *ALANINE AMINOTRANSFERASE* (*AlaAT)* and *ASPARTATE AMINO-TRANSFERASE* (*AspAT)* and a mitochondrial *MALATE DEHYDROGENASE* (*MDH)*, which are essential for the conversion of transfer acids, were more abundant in the terrestrial culms (Figure 6A). In accordance with this C_4 photosynthesis profile, the carbon isotope ratio ($\boldsymbol{\delta}^{13C}$) with -15.76 ‰ was comparable to that of other C_4 plants (Figure 6B; Cernusak et al. 2013).

The aquatically grown culms had decreased C_4 cycle enzyme expression, though the C_4 photosynthesis signature was still higher as compared to typical C_3 plants. In the aquatic culms Calvin-Bassham-Benson cycle (CBBC) related transcripts were up-regulated (Fisher's Exact BY corrected *P*-value 1.09E-14; Figure 5A; Supplemental Dataset 6). Furthermore, the small *RUBISCO* subunit (Fisher's exact test BY corrected *P*-value 3.80E-06; Supplemental Dataset 6) and the *RUBISCO ACTIVASE* (RCA; Fisher's exact test BY corrected *P*-value 0.026; Figure 5A; Supplemental Dataset 6) were significantly enriched. At the same time transcripts related to photorespiration were strongly up-regulated (Fisher's exact test BY corrected *P*-value 7.10E-08; Figure 5A; Supplemental Dataset 6). Especially, *SERINE HYDROXYLMETHYLTRANSFERASE* (*SHM)* and *GLYCINE DECARBOXYLASE COMPLEX* (*GDC)* subunits were significantly up-regulated in the aquatic culms (Figure 7; Supplemental Dataset 6). Also the *GLYCOLATE OXIDASE* (*GOX)* was up-regulated in the aquatic transcriptome (BH corrected *P*-value 0.001; Dataset 1). Complementing the strong photorespiratory signature, transcripts related to refixation of photorespiratory ammonia via glutamine/glutamate synthesis were up-regulated in the aquatic as compared to the terrestrial culms (Figure 7; Supplemental Dataset 6 and 7). Reflecting this less pronounced C_4 signature, the carbon isotope ratio of the aquatic culms with -19 ‰ (δ^{13C} value) showed stronger discrimination against ^{13}C, which is closer to the range of C_3 plants (Figure 6B; Cernusak et al. 2013).

The C_4 cycle transcripts in the terrestrial culms showed enhanced plasticity depending on water availability and drought stress

During its lifecycle, *Eleocharis* plants can be subjected to great changes in its habitat of fresh water streams and ponds, including episodes of flooding and drying (Ueno 2001). Consequently, these sedges have evolved the ability to adjust their phenotype quickly to changing environments. Signals of adjustments and transcriptomic plasticity were detected in by comparative transcriptomic analysis. Genes that were involved in habitat acclimatization, showed a high degree of variation between the replicates of the same growth condition (Figure 2C).

To independently corroborate this finding, we tested the variability of C_4 cycle genes under different degrees of drought. To this end, we grew *E. retroflexa* plants on soil and provided them with decreasing amounts of water per group for two weeks (Group 1: control every day 250ml water; Group2: every two days 250ml water; Group3: every four days 150ml water; Group4: no water for 14 days). From these plants transcripts of core C_4 cycle enzymes (*NAD-ME*, *PPDK*, *PEPC*) were measured via qRT-PCR (Figure 8A-C). Moreover, PEPC, NAD-ME, PEPCK, AspAT and AlaAT enzyme activities were determined by coupled photometric assays (Figure 8D). Based on transcriptional activity, reducing the water amount from group1 to group2 had no significant effect on gene expression (Figure 8A-C). However, limiting the water availability to watering every fourth day (group3) caused a significant increase in the expression of *NAD-ME* (3-fold), *PPDK* (7-fold) and *PEPC* (13-fold) between group 1 and 3 (*P*-value<0.001; Figure 8A-C). The enzyme assays showed a trend towards increasing PEPC and NAD-ME activity during drought, however, the magnitude of change was much lower and the changes were not significant between group1 and group3 (Figure 8D). Thus, enzyme activity remains more stable, whereas *E. retroflexa* reacts strongly on transcriptional level to environmental stimuli.

To investigate what might be controlling this drought response, we took a closer look at the comparative transcriptomes. Abscisic acid (ABA) metabolism (BY corrected *P*-value 0.0041) and synthesis (BY corrected *P*-value 0.0038) were enriched in the terrestrial culms (Supplemental Dataset 6). In a related species, *E. vivipara*, changes in C_4 cycle enzymes are tied to changes in ABA concentration (Agarie et al. 2002; Ueno 2001; Ueno et al. 1988) Transcripts related to the biosynthesis or degradation of other phytohormones were not detected as differentially expressed (Supplemental Dataset 6; Supplemental Figure 7). On the level of transcriptional regulators, 101 transcription factors (TFs) were differentially transcribed between terrestrial (26 TFs up-regulated) and aquatic culms (75 TFs up-regulated; Dataset 1). Interestingly, we detected significant changes related to Histone modification (BY corrected *P*-value 0.00059), DNA methyltransferases (DNMT) and (*DMT7*; *DNMT2*; *MET1*; *DRM1*; *CMT1*; Wilcoxon rank test BY corrected *P*-value 0.0137). Furthermore, *ALIFIN-LIKE 1* (*AL1*) transcriptional regulators (*AL1*; *AL3*; *AL5*; *AL6*; *AL7*; Wilcoxon rank test BY corrected *P*-Value 0.012) were enriched in the aquatic culms (Supplemental Figure 7).

Discussion

E. retroflexa has been described as NAD-ME C_4 photosynthesis performing sedge under terrestrial conditions and as a C_4–like plant when submerged in water (Ueno and Wakayama 2004; Uchino et al., 1995; Ueno 2004). In general, C_4 plants have been proposed to display less plasticity in the range of growth habitats and phenotypic plasticity (Sage and McKown 2006). To determine, which transcriptional programs enable *E. retroflexa* to acclimatize to terrestrial and aquatic lifestyle so rapidly and successfully, meanwhile performing C_4/C_4-like photosynthesis, we generated two comparative transcriptomes of aquatically and terrestrially grown *E. retroflexa* culms. This data was contextualized with physiological and anatomical parameters, and experiments assessing the adaptability of the C_4 cycle under drought stress. A phylogenetic analysis based on ITS sequences revealed that *E. retroflexa* is much more closely related to *E. baldwinii* than *E. vivipara*. Amongst these previously described C_4 species *E. retroflexa* and *E. baldwinii* have a stronger C_4 signature compared to *E. vivipara*, which has been suggested to have evolved C_4 photosynthetic traits more recently (Supplemental Figure 8; Ueno 2001). It is unclear, whether *E. retroflexa* is still evolving towards full C_4-ness or if the display of C_3-C_4 intermediate traits is a reversion from C_4 photosynthesis for better adaption under water.

The *E. retroflexa* unigene database provides a base for further molecular studies

To date no Cyperaceae genome is available and transcriptomes have only been published for *Eleocharis baldwinii* recently (Chen et al. 2014). With this study we provide a reference database of 27,021 unigenes for *E. retroflexa* that covers most of all core plant pathways and most cellular processes and regulatory pathways, as represented by the KEGG database (Moriya et al., 2007; Supplemental Dataset 1, 2 and 3). With an N50 of 1,199 bases and average contig length of 789 bases our filtered unigene databse is in the range of other *de novo* assembled reference transcriptomes, such as radish (*Raphanus sativus*; Wang et al. 2013), scarlet sage (*Salvia splendens*; Ge et al. 2014), *Megathrysus maximus* and *Dichantelium clandestinum* (Bräutigam et al. 2014). The assembled contigs provide the basis for designing primer-sets for qRT-PCR, indicating that the database represents the *E. retroflexa* transcripts to a replicable degree (Figure 8A-C).

For differential transcriptome analysis we opted for cross-species mapping of reads rather than mapping the reads to the unigene database as *de-novo* contig assemblies of short RNA-seq reads still suffer from several shortcomings. A major issue is the occurrence of redundant contigs representing one gene locus (Papanicolaou et al. 2009), which we also observed in our assembly (Figure 1B). This artificial inflation of contig numbers occurs particularly frequent for highly expressed genes, as well as genes that contain highly conserved sequence motifs of large gene families, such as transcriptional

regulators (Figure 1B). These redundant contigs display slight differences between each other, due to alternative splicing, sequencing errors, or single nucleotide polymorphisms (SNPs) between alleles (Papanicolaou et al. 2009). Another disadvantage arises from lowly expressed genes with subsequent low read coverage or assembly errors leading to either fragmented contig or absence of transcripts (Martin and Wang 2011; Schliesky et al. 2012).

Cross-species mapping introduces less bias in expression level dynamics than mapping to contigs

Mapping to related reference genomes can pose challenges for subsequent data analysis. For example, species-specific genes cannot be mapped because they are not represented in the reference. In addition, different genes might display different rates of sequence divergence due to different evolutionary rates, which might lead to a bias in quantifying gene expression levels. However, the advantages of cross-species mapping outweigh the possible shortcomings of cross-reference mapping and this method has been successfully used for transcriptome analyses of other non-model species with no available reference genome (Bräutigam et al. 2010; Bräutigam et al. 2014; Gowik et al. 2011; Külahoglu et al. 2014).

To evaluate whether cross-species mapping itself leads to any systematic bias on estimating expression levels, we visualized the fraction of reads mapping to higher and lower expressed sequences in the recently sequenced species *Tarenaya hassleriana* to its own genome (Cheng et al. 2013) versus the genome of the model species *A. thaliana* (Lamesch et al. 2012; Supplemental Figure 2B). This comparison shows that, in contrast to contig mapping, mapping to a cross-species reference does not cause visual alterations in the mapping dynamics.

Transcriptional changes between the terrestrial and the aquatic culms are closely related to the habitat switch and photosynthetic mode

To put the amount of changes seen between the aquatic and terrestrial culms (mean r=0.86; Supplemental Table 3) in context, the correlation between the two transcriptomes is compared with published expression data from two closely related C_4 (*Gynandropsis gynandra*) and C_3 (*Tarenaya hassleriana*) Cleomaceae species (Külahoglu et al. 2014). The comparison of aquatic and terrestrial *E. retroflexa* culms shows a slightly higher similarity based on Pearson's correlation than between the mature leaves of *T. hassleriana* (C_3) and *G. gynandra* (C_4; mean r=0.80; Supplemental Table 3). At the same time, comparing the transcriptomes of a leaf derived tissue, such as petals against mature leaf of *T. hassleriana* (mean r= 0.11), or root against mature leaf (mean r=0.09; Supplemental Table 3), reveals that the adaptation of *E. retroflexa* culms to different habitats is changing the

transcriptomes general dynamic only to a minor degree. Thus, we conclude, that the changes monitored between the terrestrial and the aquatic culms should reflect the habitat switch and photosynthetic mode of the culm tissues.

There is a high similarity between the aquatic and the terrestrial culm transcriptomes based on PCA and HCL with around 8% (1,356 genes) of all transcripts being detected as significantly different (Figure 2A-B; Supplemental Figure 3). The number of altered transcripts between the two habitats is comparable to the number of transcripts responding to systemic responses, pathogen or pest attack (9%; De Vos et al. 2005). Interestingly, it is higher than the number of transcript detected as significantly changed between closely related C_4 and C_3 species, even though overall gene expression patterns correlate more closely (~4% in *Cleome* and 3.4 % in *Flaveria*; Bräutigam et al. 2010; Gowik et al. 2011).

The transcriptomes of *E. retroflexa* show high plasticity and variability in transcripts linked to the habitat acclimatization

Biological variation of transcript abundances between each of the replicates of the same habitat displayed a clearly distinct variation, as judged on the basis of PCA and Pearson's correlation values (Figure 2A-B; Supplemental Table 3). The transcripts displaying the strongest fluctuation between biological replicates are identical with those that are significantly differentially regulated between the aquatic and terrestrial habitat (Figure 2C). This indicates that transcriptional variation between the replicates is not arbitrary or due to experimental error, but it is rather associated with transcriptional fine-tuning of culm acclimatization.

Transcript abundance of enzymes associated with C_4 photosynthesis (*PPDK*, *PEPC*, *NAD-ME*) increases under water deprivation (Figure 8A-C), possibly induced through the hormone abscisic acid (ABA). These results indicate that microenvironmental cues can significantly affect the transcriptional program of *E. retroflexa*. For *E. vivipara*, a related *Eleocharis* species, it has been shown that ABA is able to induce C_4-ness under submerged conditions (Agarie et al. 2002; Ueno et al. 1988). Similarly, ABA signaling has been reported to induce CAM photosynthesis in some facultative CAM plants (Chu et al. 1990; McElwain et al. 1992). The hormones ABA and ethylene have been connected to aquatic leaf formation and its regulation in heterophyllous amphibious plant species (Kuwabara et al. 2001; Minorsky 2003). In the terrestrial transcriptome ABA metabolism is significantly up-regulated (Supplemental Dataset 6), with no trace of other plant hormone circuits being significantly altered under aquatic or terrestrial conditions (Supplemental Figure 7). For *E. vivipara* it has been reported that application of exogenous gibberilic acid to terrestrial culms can trigger submerged (C_3) culm anatomy featuring small BSCs and the absence of stomata (Ueno 2001). However, no enhanced

gibberilic acid signaling could be detected in the aquatic transcriptomes (Supplemental Figure 7), indicating that on transcriptional level different regulatory mechanisms may play a role in *E. retroflexa*.

Under aquatic conditions plants can suffer from hypoxia and impeded gas exchange leading to increased ethylene concentrations within the submerged plants (Bailey-Serres and Voesenek 2008; Jackson, 2008). In the aquatic *E. retroflexa* transcriptome no significant alterations related to hypoxia induced signaling pathways were detected (Lee et al. 2011; Mustroph et al. 2009), implying that *E. retroflexa* culms were well acclimatized to their aquatic habitat at the time of their harvest.

Instead we find significant changes related to histone modification and DNA methyltransferases (DNMT; Supplemental Figure 7; Supplemental Dataset 7). Five *ALIFIN-LIKE* transcriptional regulators are also enriched in the aquatic culms. Among those *AL1*, *AL5*, *AL6* and *AL7* are known to bind to di- and trimethylated histone H3 at lysine 4 (H3K4me3/2), which are markers of transcriptionally active chromatin (Lee et al. 2009). Enrichments of trimethylation of histone H3 Lys4 (H3K4me3) and acetylation of histone H3 Lys9 (H3K9ac), often used as a positive marker of histone modifications, are associated with transcriptional activity and correlate with gene activation in response to drought stress (reviewed by Kim et al. 2010). In rice, modification levels of acetylation of histone H3, dimethylation of histone H3 Lys4 (H3K4me2) and trimethylation of histone H3 Lys4 (H3K4me3) are altered on submergence-inducible genes during the process from submergence to re-aeration (Tsuji et al. 2006). There, the submergence treatments resulted in the decrease of H3K4me2 levels and increase of H3K4me3 levels on the 5′- and 3′-coding regions of submergence inducible genes alcohol dehydrogenase1 (*ADH1*) and pyruvate decarboxylase1 (*PDC1*) genes (Tsuji et al. 2006).

In summary, up-regulation of transcripts encoding histone-modifying enzymes could play a role in altering the overall expression profile in the submersed culms. Interestingly, submersion of *E. retroflexa* culms does not leave traces of hypoxia-stress induced signaling in the surveyed transcriptome.

E. retroflexa culms reflect acclimatization to the habitat by changes in culm structure and photosynthesis

Earlier studies by Ueno and colleagues showed that *E. retroflexa* plants develop new adapted photosynthetic culms under water, which rapidly dry out when the plants are transferred to soil (Ueno 2001; Ueno and Wakayama 2004). When grown under water, *E. retroflexa* culms grow fast; however, they never lift beyond the water surface (Supplemental Figure 4B), which is similar to *E. vivipara's* growth habitus (Ueno 2001). This could be connected with a decreased investment in genes related to

phenolic compounds, phenylpropanoid and lignin biosynthesis in aquatic culms (Figure 3A). Characteristically, submerged leaves have two main strategies to survive in the wet habitat, by either growing out of the water by stem elongation or the development of aquatically accustomed leaves (Mommer and Visser 2005). Clearly, the latter is true for *E. retroflexa*. Comparison of cross sections of aquatic and terrestrial culms, revealed decreased auto-fluorescence of phenolic compounds indicating the presence of lignin (Figure 4B). Less investment in vascular bundles and lignin are a characteristic for aquatic plants (Sculthorpe 1967). Aquatic plants apparently need less cell wall reinforcement in the water, since a cell wall constrains the rate and direction of turgor-driven cell growth (Bailey-Serres and Voesenek 2008). Also evident in the transcriptome is the significantly higher fold-change in transcript levels related to cell wall modification and enlargement, such as various expansins in the terrestrial culms (Figure 3A, Supplemental Dataset 7). This seems to be concomitant with the observed enlarged BSCs in the terrestrial culms (Figure 4; Ueno and Wakayama 2004). Besides the decreased need for leaf rigidity under water, the aquatic environment poses unique challenges for photosynthesis by low light availability and a shift in light spectrum (Kirk 1994; Pedersen et al. 2013). We find that *E. retroflexa* overcomes these challenges by a 10% higher investment in transcripts related to light reactions and photosystems (Supplemental Figure 6; Figure 5A). The aquatically adapted plant *Rumex palustris* shows lower PSII abundance after acclimatization to submergence compared to terrestrially acclimation (Mommer et al. 2005b). In submerged *E. retroflexa* transcriptomes both photosystems are increased (Figure 5A). When submerged, *E. retroflexa* culms have a higher demand of light absorption, which is structurally supported by higher abundance of transcripts related to galactolipid biosynthesis and transcripts needed for thylakoid membrane assembly (Figure 5A). The higher transcript abundances portioned into both photosystem polypeptide subunits, light harvesting complexes and in chlorophyll and carotenoid biosynthesis, indicate that aquatic culms have to adjust their photosynthetic apparatus as an adaption to lower light intensities and a change in light spectrum under water (Holmes and Klein 1987; Sand-Jensen 1989). Interestingly, during shade avoidance plants display higher chlorophyll per dry weight (Bailey et al. 2004), similar to what is observed in submerged *E. retroflexa* for chlorophyll content (Figure 5B). Hence the adaption of the light harvesting machinery of *E. retroflexa* might derive from a shade avoidance response (SAR) as it has been suggested for other submerged growing plant species (Boeger and Poulson 2003; Frost-Christensen and Sand-Jensen 1995; Mommer et al. 2005a).

On the regulatory level the *GOLDEN-2-LIKE 1* (*GLK1*) transcription factor is significantly (three-fold) up-regulated in the aquatic culms (150/50 RPM). Prior studies in the C_4 plant *Zea mays* showed that the *GOLDEN-2* gene is exclusively expressed in the BSCs, whereas *GLK1* is only expressed in MCs (Langdale and Kidner 1994; Rossini et al. 2001). In both cell types these

transcription factors are important for chloroplast biogenesis (Rossini et al. 2001). The results presented here for aquatic culms are consistent with the described function of GLK1 acting as nuclear regulator of photosynthetic capacity (Waters et al. 2009), especially under submerged conditions when MCs are being enlarged and accumulate more chloroplasts (Figure 4A).

E. retroflexa culms display a change in its C_4 photosynthesis profile depending on environmental cues

E. retroflexa has been described as a NAD-ME subtype C_4-like photosynthesis performing species under aquatic and terrestrial conditions, though the C_4 cycle enzymes (PPDK, PEPC and NAD-ME) show slightly lower protein abundance in the aquatic form (Ueno et al., 2004; Figure 6A). In both transcriptomes and in the enzyme activity assays we found evidence for the presence of two decarboxylating enzymes –NAD-ME and PEPCK (Figure 8D), as it has been described for *G. gynandra* and *M. maximum* (Bräutigam et al. 2014; Sommer et al. 2012).

During submergence, photosynthesis rates drop in non-adapted terrestrial species, due to the increased gas diffusion resistance, restricted access to light and biochemical limitations (Centritto et al. 2003; Long and Bernacchi 2003). Aquatic acclimated culms are thinner and have reduced cuticles to decrease the internal diffusion path for CO_2 to the chloroplasts (Maberly and Madsen 2002; Madsen and Sandjensen 1991). Aquatic *E. retroflexa* culms visually appear to be thinner, have less dry matter (Figure 3B) and their transcriptional investment in cuticular waxes is down-regulated (Figure 3A). The reduction of cuticle thickness has been reported to led to a reduction of the gas diffusion resistance in aquatic plants (Frost-Christensen et al. 2003). Typical C_4 architecture is dissolved in aquatic culms: the MCs adjacent to epidermis are massively enlarged with high chloroplast content in aquatic culms (Figure 4A), as it has been reported for other aquatically adapted plants (Mommer and Visser 2005). For reducing the diffusion path length, the chloroplasts are present in all epidermal and sub-epidermal cells and positioned towards the exterior of the cells (Mommer et al. 2005b). All these monitored acclimations of aquatic *E. retroflexa* culms support the hypothesis that CO_2 directly enters the MCs of the aquatic leaves via diffusion through the epidermis and not via stomata (Mommer et al. 2005b).

Another mechanism of aquatic plants for reducing gas diffusion resistance is the conversion of CO_2 to HCO_3^- for higher solubility catalyzed by carbonic anhydrases (CA; reviewed by Pedersen et al. 2013). In *E. retroflexa* one *CARBONIC ANHYDRASE (CA2)* appears to be recruited to the C_4 cycle by its up-regulation in the terrestrial culms, whereas in the aquatic culms the BETA *CARBONIC ANHYDRASE 5* is 1.5 fold higher accumulated (Dataset 1; BH corrected *P*-value 0.0003).

Per definition C_4-like species display higher C_4 cycle activities than C_3-C_4 intermediate, but lack complete BSC compartmentation of RuBisCO (Edwards and Ku 1987). It has been postulated earlier, that *E. retroflexa* culms maintain a C_4-like profile when submerged based on C_4 cycle enzymes and

RuBisCO protein immuno-localizations (Ueno and Wakayama 2004). Transcriptome analysis revealed, that the terrestrial culms have a stronger C_4 cycle signature than the aquatic culms (Figure 6A). The small subunit of the *RUBISCO* is two-fold more highly expressed in the aquatic culms (Dataset 1) and the large RuBisCO subunit is present in BSCs as well as MCs (Ueno and Wakayama 2004). However, the aquatic culms show an untypically high expression of transcripts related to photorespiration for either aquatically adapted (Mommer et al. 2006) or C_4-like plants (Mallmann et al. 2014; Figure 5A). Especially, *SHM* and *GDC* subunits are significantly up-regulated in the aquatic culms (Figure 7). Typically, underwater photosynthesis in non-acclimated terrestrial plants is characterized by high photorespiratory rates, as reduced gas diffusion rates under water will lead to relatively low internal CO_2 concentrations compared with the internal oxygen concentrations in the presence of light (Jahnke et al. 1991; Maberly and Spence 1989). Similar conditions can occur for aquatic adapted plants, when they grow as dense a canopy, which leads CO_2 depletion and O_2 supersaturation of the water during day time (Keeley 1999). In the genus *Flaveria* C_3-C_4 intermediate species have been reported to display a similar photorespiratory signature as C_3 species (Mallmann et al. 2014), while maintaining the C_4 cycle in parallel. In evolutionary terms, the establishment of a photorespiratory CO_2 pump –also termed as C_2 photosynthesis- is thought to be a necessary step towards C_4-ness (Gowik et al. 2011; Heckmann et al. 2013; Mallmann et al. 2014; Sage 2004; Schulze et al. 2013).

In aquatic *E. retroflexa* culms this high photorespiratory signature might stem from two factors: (i) abolishment of the strict RuBisCO compartmentalization to the BSCs in the aquatic culms as seen in the terrestrial culms (Ueno and Wakayama 2004) and (ii) dense vegetation of E. retroflexa plants leading to CO_2 depletion under water.

A possible explanation for the photorespiratory signature not being associates with an apparent fitness penalty in aquatic *E. retroflexa* culms could be that the photorespiratory cycle is used for an efficient CO_2 re-fixation and balancing O_2 and CO_2 availability at the site of RuBisCO, which arises from higher-diffusion resistance for CO_2 uptake and continuous photosynthesis action. Thus, these plants might complement thereby the C_4 cycle possibly by using photorespiration for cycling each intercellular CO_2 until it is fixed in form of carbon compounds, as it is known from C_3-C_4 intermediate species.

Besides the genus *Eleocharis* the monocotyledonous Orcuttiea family has amphibious C_4 species (Keeley 1998). When grown on soil these species perform C_4 photosynthesis and submerged they switch to C_4-like photosynthesis without classic Kranz anatomy (Keeley 1998). Single-cell C_4 photosynthesis has been also found in facultative aquatic species, e.g. *Hydrilla* and *Egeria* under limited CO_2 availability and warm water temperatures (Bowes et al. 2002; Casati et al. 2000; Reiskind et al. 1997).

So far no examples of classic two-cell BSC/MC C_4 photosynthesis have been discovered for aquatic plant species. With the variety of mechanisms evolved to circumvent the gas diffuse

resistance and optimize CO_2 fixation, one may wonder whether performing two-cell C_4 photosynthesis is actually feasible under water.

Conclusions

In this study, we present an in depth analysis of *E. retroflexa* transcriptional acclimatization to terrestrial and aquatic habitats. The assembly of the transcriptomes provides a unigene database for further molecular studies. The transcriptomes of the terrestrial and aquatic culms now enabled a detailed analysis of *E. retroflexa*'s full C_4 cycle, carbon concentrating mechanism and metabolism under different growth condition. The aquatic *E. retroflexa* transcriptome reflects many traits known for other heterophyllous aquatic plant species. *E. retroflexa* is surprisingly flexible in its usage of the C_4 cycle and reacts fast to micro-environmental changes, such as water deprivation. While classic Kranz anatomy is lost under water, *E retroflexa* possibly uses a C_2-like photorespiratory cycle to supplement the C_4 cycle as seen in C_3-C_4 intermediate plant species.

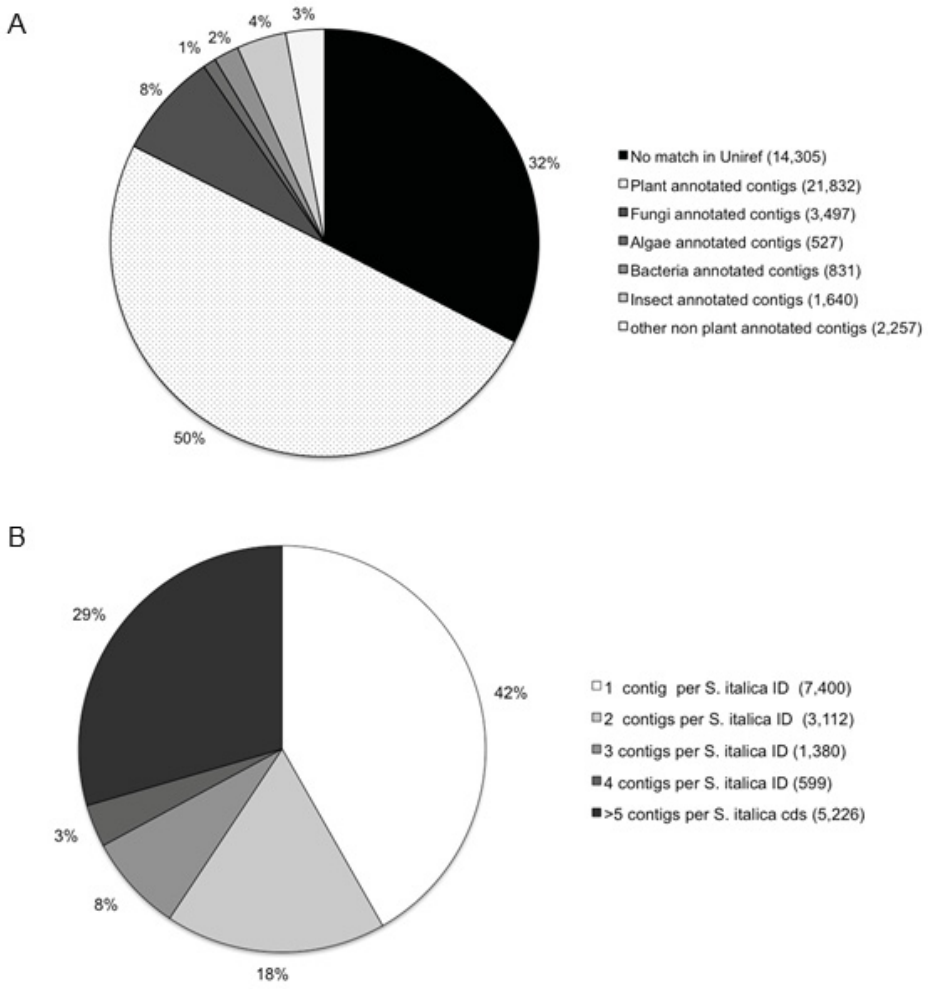


Figure 1. Annotation of *E. retroflexa* contigs by blastx searches
(A) Annotation of contigs against UniRef 100. Distribution of contigs annotated by UniRef100 falling into major species categories of plant, bacteria, algae, fungi and other non-plant annotated contigs. Total contig number is indicated in parentheses.
(B) Annotation of contigs against *S. italica* gene models. Percentage of contig number per predicted primary cds showing redundancy in assembled contigs. Number of best matching contigs per predicted *S. italica* id is indicated in parentheses.

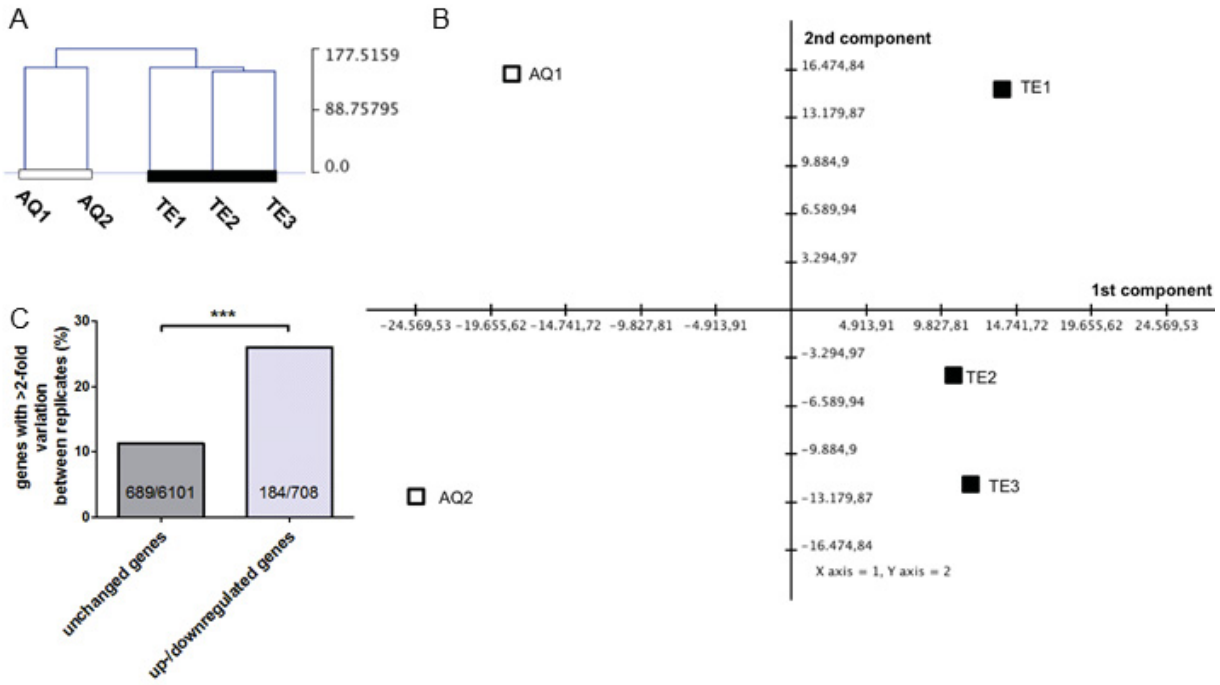


Figure 2. Transcriptome dynamics and variance in between samples and habitat.

(A) Hierarchical clustering of all sequenced *E. retroflexa* samples. *E. retroflexa* transcriptomes (>1 RPM filtered) were clustered after normalization with Euclidean distance and average linkage. AQ, submerse aquatic; TE, terrestrial.

(B) Principle component analysis between aquatic and terrestrial grown *E. retroflexa* transcriptomes.

Plot shows all sequenced samples from aquatically (white) and terrestrially (black) grown *E. retroflexa* (n=3; RPM). First component (x-axis) separates samples by habitat (36%) of all data variability, and second component (y-axis) describes biological sample variability (21%) within each growth condition.

(C) Variance plot between replicates and significant changed transcripts.

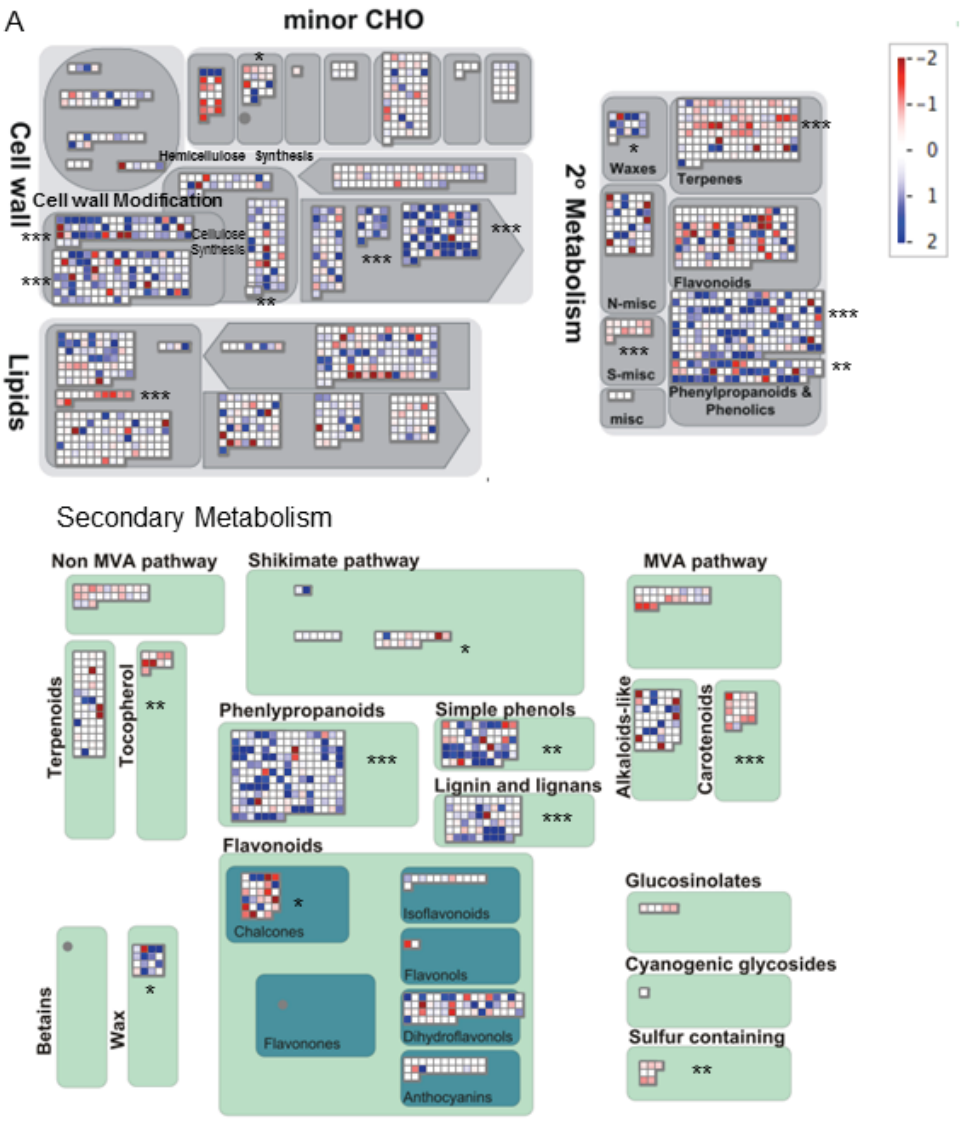


Figure 3. Structural differences between aquatically and terrestrially grown *E. retroflexa* culms.

(A) Overview of secondary and carbon metabolism gene expression patterns (Mapman, TAIR10) in *E. retroflexa* culms. Heatmaps depict transcriptional log2 fold-changes aquatic versus terrestrial RPM. Red (ratio<0) represent an decrease of gene expression in aquatic and blue (ratio>0) an increase of transcript accumulation in terrestrial culms. Asterisks indicate significant fold-changes, that were calculated by Wilcoxon Rank Test and P-values (* P value<0.05; ** P-value<0.01; ***P-value<0.001) were Benjamini-Yekutieli FDR corrected. Heatmaps were generated with Mapman tool V3.5 (Usadel et al., 2006). **(B)** Comparison of water content and biomass between terrestrially (black) and aquatically grown culms as ratio dry weight (DW) against fresh weight (FW) in percent. n= 3 biological replicates;

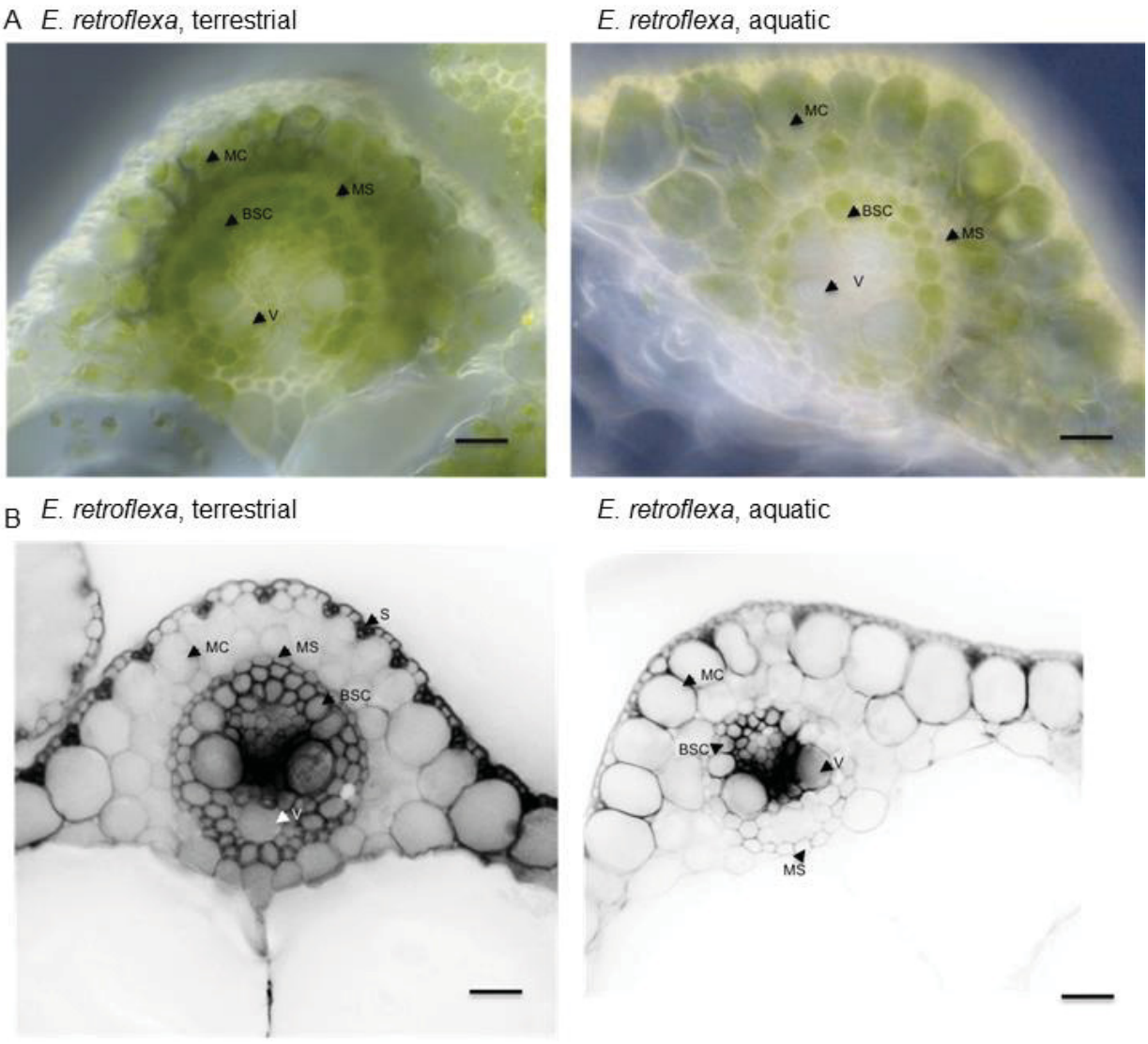


Figure 4. Culm anatomy of mature *E. retroflexa* plants grown under terrestrially (left) and aquatically (right) conditions. (A) Microscopic images of *E. retroflexa* cross sections grown on soil and under water. **(B)** Autofluorescence microscopic images of *E. retroflexa* cross-sected culms of grown on soil and under water. Scale bar: 20µm. Cell types are indicated by closed arrows. BSC: bundle sheath cell; MC: mesophyll cell; MS: mestome sheath; V: vein.

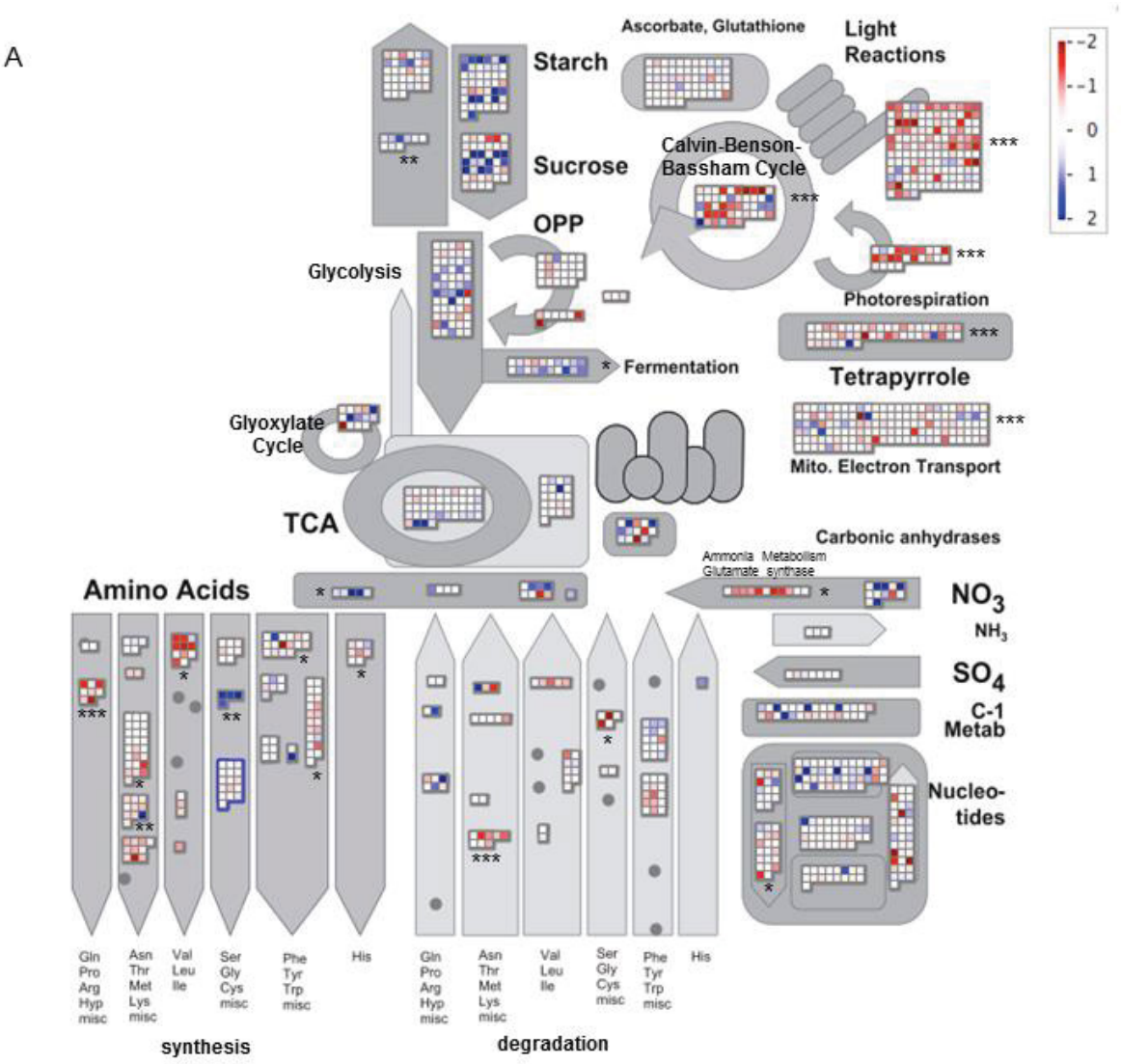


Figure 5. Transcriptional and physiological differences between aquatically and terrestrially grown *E. retroflexa* culms.
(A) Overview of central metabolism gene expression patterns (Mapman, TAIR10) in *E. retroflexa* culms. Heatmaps depict transcription log2 fold-changes aquatic versus terrestrial RP Red (ratio<0) represent an decrease of gene expression in aquatic and blue (ratio>0) an increase of transcript accumulation in terrestria culms. Asterisks indicate significant fold-change These were calculated by Wilcoxon Rank Test P-values (* P-value<0.05; ** P-value<0.01; ***P value<0.001) were Benjamini-Yekutieli FDR corrected. Heatmaps were generated with Map tool V3.5 (Usadel et al., 2006)
(B) Total chlorophyll content in terrestrially (blac and aquatically grown culms as µg per mg dry

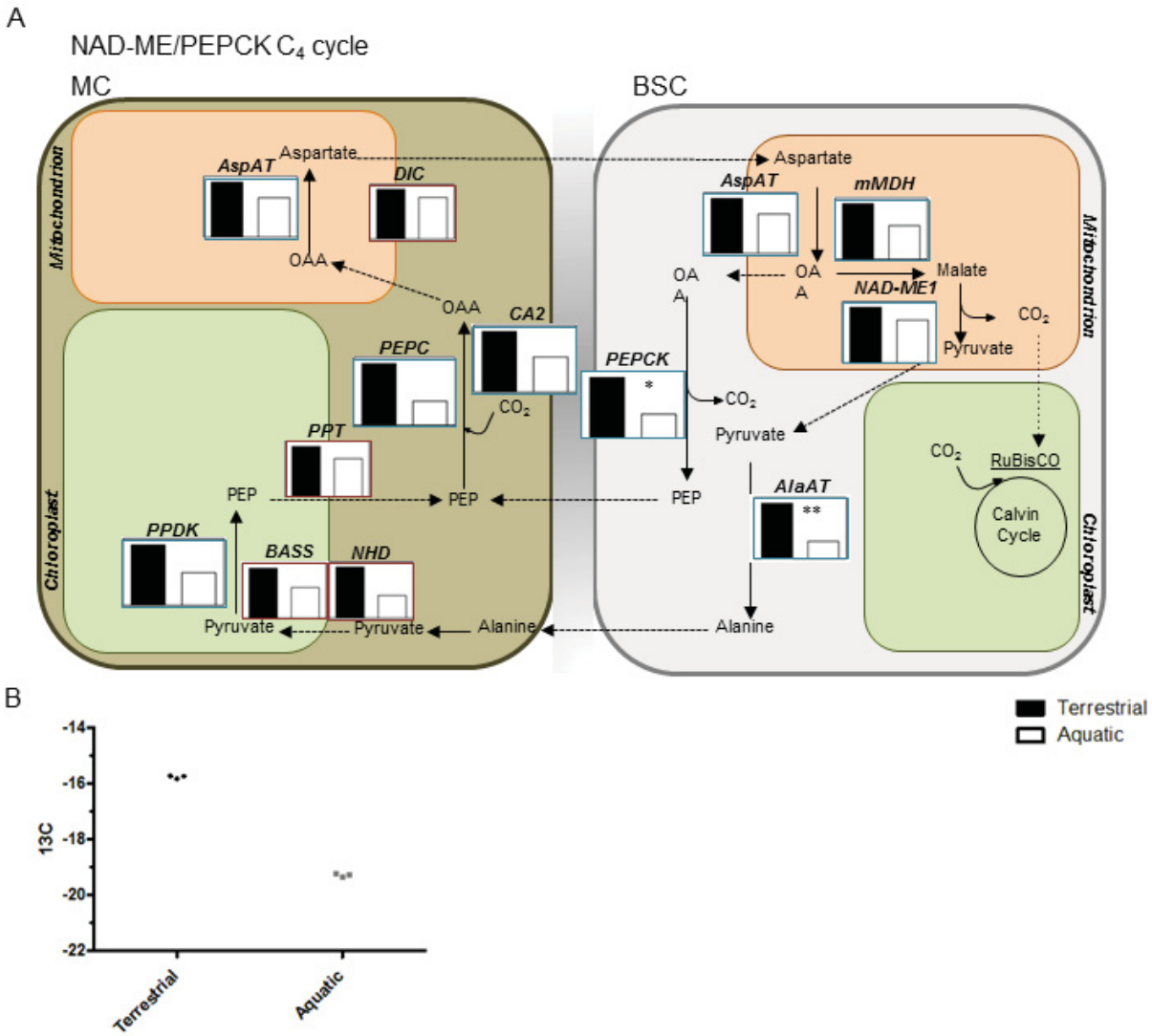


Figure 6. The C_4 cycle is altered between terrestrially and aquatically grown *E. retroflexa* culms.

(A) Schematic and simplified overview of the NAD-ME/PEPCK C_4 cycle known for C_4 plants (adapted from Sommer et al., 2012). Relative transcript abundances between terrestrial (black) and aquatic (white) transcriptomes are shown in small insets and were normalized by setting the highest expressed condition to 1 for each gene. Asterisks denote significant expression changes between aquatic and terrestrial samples (Edge R; FDR BH corrected P-values). ** P-value<0.01; * P-value<0.05.

Localization of C_4 enzymes in *E. retroflexa* is assumed from literature (Ueno et al., 2004). Red boxes indicate relevant C_4 cycle transporter and blue boxes soluble C_4 cycle enzymes. ***PEPC***: *PHOSPHOENOLPYRUVATE CARBOXYLASE*; ***CA2***: *CARBONIC ANHYDRASE2*; ***DIC***: *DICARBOXYLATE CARRIER*; ***AspAT***: *ASPARTATE AMINOTRANSFERASE*; ***mMDH***: *mitochondrial MALATE DEHYDROGENASE*; ***NAD-ME1***: *NAD-dependent MALIC ENZYME1*; ***AlaAT***: *ALANINE AMINOTRANSFERASE*; ***PEPCK***: *PHOSPHO*ENOL*PYRUVATE CARBOXYKINASE*; ***BASS***: *BILE ACID:SODIUM SYMPORTER*; ***NHD***: *SODIUM:HYDROGEN ANTIPORTER*; ***PPDK***: *PYRUVATE ORTHOPHOSPHATE DIKINASE*; ***PPT***: *PHOSPHATE/PHOSPHO*ENOL*PYRUVATE TRANSLOCATOR* **(B)** 13C/12C isotope ratio of terrestrial (black) and aquatic (grey) culms; n=3.

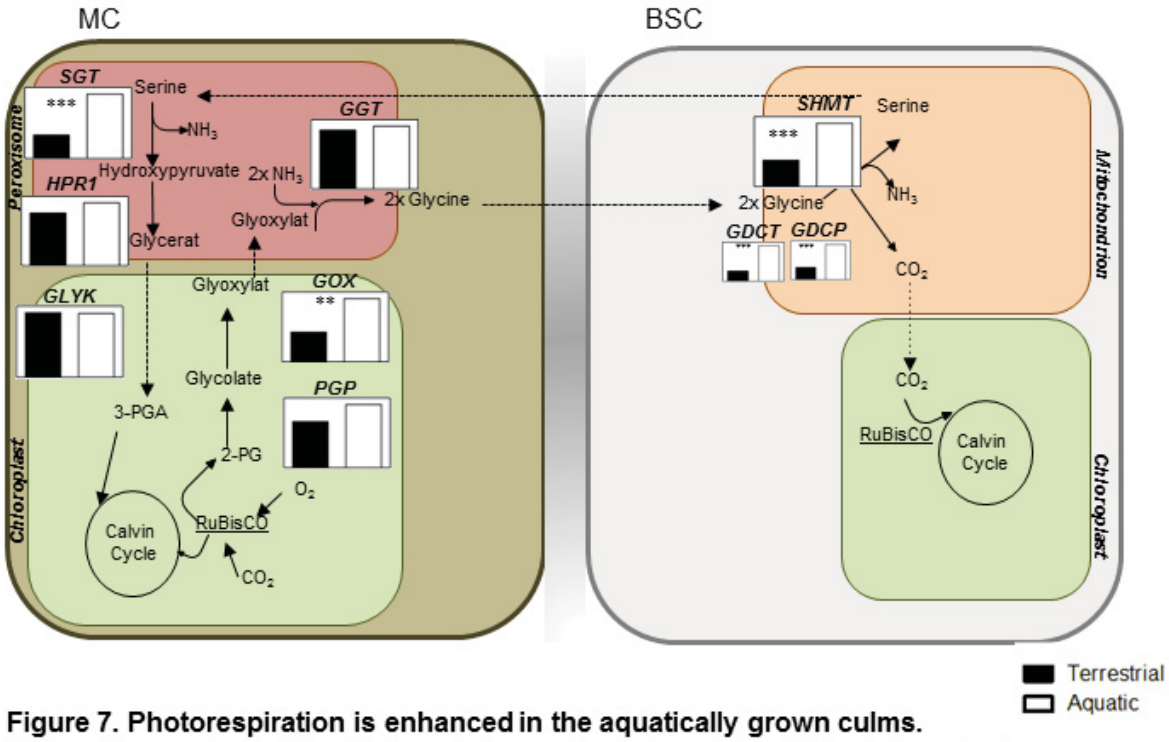


Figure 7. Photorespiration is enhanced in the aquatically grown culms.
Schematic over view of the photorespiratory pathway known for C_4/C_3 photosynthesis intermediate plants (adapted from Gowik et al. 2012). Relative transcript abundances between terrestrial (black) and aquatic (white) transcriptomes are shown in small insets and were normalized by setting the highest expressed condition to 1 for each gene. Asterisks denote significant expression changes between aquatic and terrestrial samples (Edge R; FDR BH corrected P-value). ** P-value<0.01; *** P-value<0.001.
Localization of photorespiratory enzymes is assumed from literature (cite).
PGP*: *2-PHOSPHOGLYCOLATE PHOSPHATASE*; *GOX*: *GLYCOLATE OXIDASE*; *GGT*: *GLUTAMATE:GLYOXYLATE OXIDASE*; *GDC*: *GLYCINE DECARBOXYLASE*; *SHM*: *SERINE HYDROXYLMETHYL TRANSFERASE*; *SGT*: *SERINE:GLYOXYLATE AMINOTRANSFERASE*; *HPR1*: *NADH-dependent HYDROXYPYRUVATE REDUCTASE*; *GLYK*: *GLYCERATE KINASE

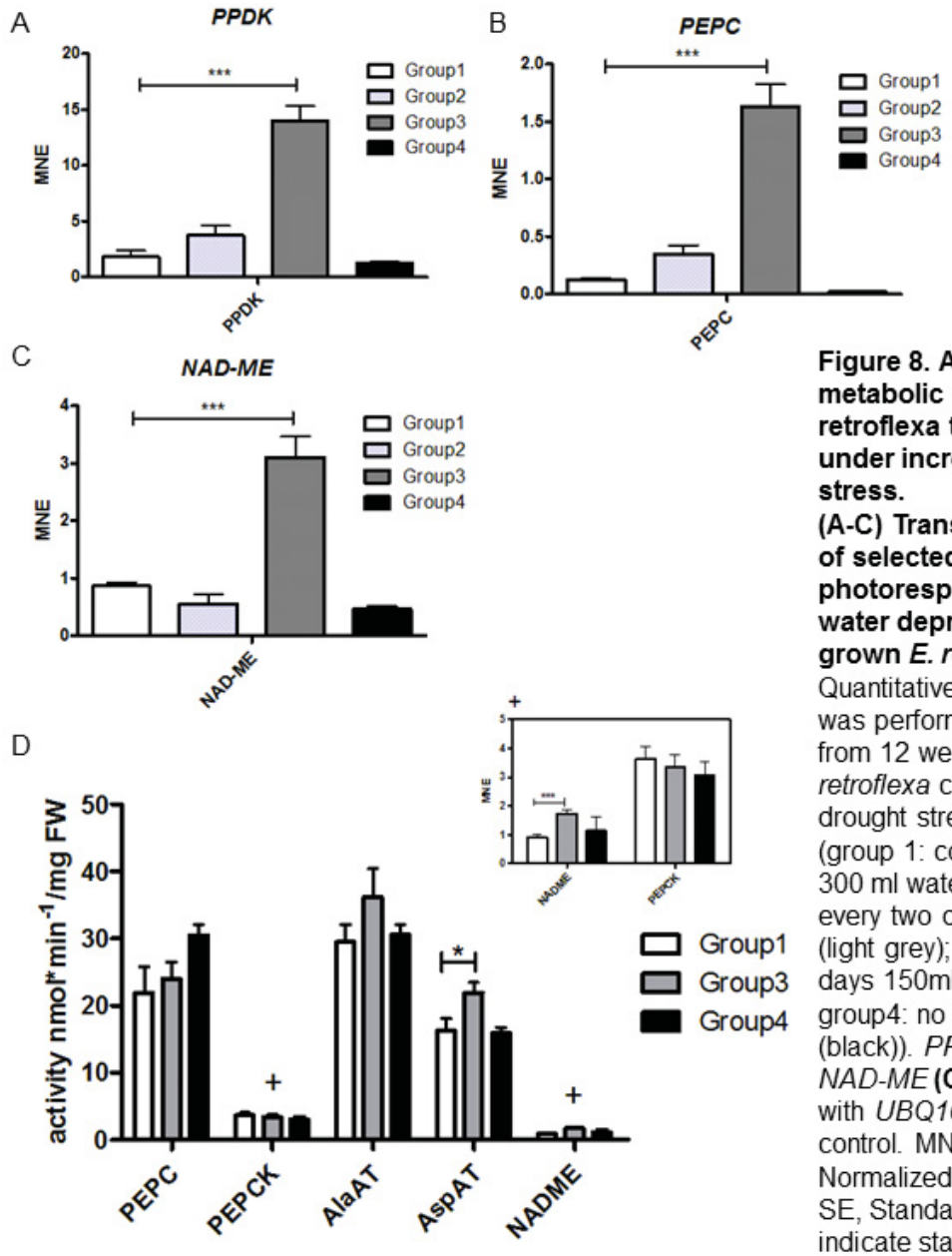


Figure 8. Analysis of metabolic plasticity of E. retroflexa terrestrial culms under increasing drought stress.
(A-C) Transcriptional pattern of selected C4 and photorespiratory genes in water deprived terrestrial grown *E. retroflexa*.
Quantitative real-time PCR was performed on samples from 12 week old terrestrial *E. retroflexa* culms subjected to drought stress for 14 days (group 1: control, every day 300 ml water (white); group2: every two days 250ml water (light grey); group3: every four days 150ml water (dark grey); group4: no water for 14 days (black)). *PPDK* **(A)**, *PEPC* **(B)**, *NAD-ME* **(C)** were normalized with *UBQ10* as housekeeping control. MNE: Mean Normalized Expression; n=3 ± SE, Standard Error. Asterisks indicate statistically significant differences between control and group 3 (***P-Value<0.001).

(D) Enzyme activity measurement of soluble C_4 cycle enzymes in water deprived terrestrial grown *E. retroflexa* . Enzyme activities of PEPC, NAD-ME, PEPCK, AspAT, AlaAT were measured from 12 week old terrestrial *E. retroflexa* culms subjected to drought stress for 14 days (Group 1: control, every day 300ml water (white); every four days 150ml water (dark grey); Group4: no water for 14 days (black)). (FW: fresh weight; error bars ±SE; 3 biological replicates each with 3 technical replicates) Asterisks indicate statistically significant differences between control and group 3 (* P-values<0.01; *** P-Value<0.001). Cross indicates insets.

Tables

Table 1. Sequencing and Mapping statistics and transcriptome dynamics of *E. retroflexa* read samples aligned to *S. italica* and *E. retroflexa* reference.

Datasets

Dataset 1. Annotated transcriptome expression data (RPM) of *E. retroflexa* aquatic and terrestrial culms.

Dataset 2. CAP3 assembled filtered *E. retroflexa* unigene database.

Supplemental Material

Supplemental Datasets

Eleocharis samples	Aquatic 1	Aquatic 2	Terrestrial 1	Terrestrial 2	Terrestrial 3
raw reads	30469686	34239945	22011612	30043954	25010427
cleaned reads	28,203,253	33,838,855	21,729,249	29,391,792	24,629,693
mapped reads to *S. italica*	7,768,905	12,337,799	7,648,795	8,182,178	7,585,432
mapped reads to unigene database (>200 bases)	7,593,124	23,674,037	18,779,815	24,787,798	17,606,148
mapping efficiency to *S. italica*	28	36	35	28	31
mapping efficiency to unigenes	27	70	86	84	71
Number of *S. italica* IDs >20 RPM	4,882	5,263	5,826	5,807	5,495
Number of *S. italica* IDs > 1,000 RPM	132	143	135	136	136
Number of *S. italica* ID matching	19,298	20,248	19,893	20,041	19,814
Number of unigenes matching	34,971	38,548	38,324	38,352	38,489
S.italica IDs covered by reads (%)	54.4	57.1	56.1	56.5	55.9
E. retroflexa unigenes covered by reads (%)	79.8	88.0	87.5	87.5	87.8
Transcript number >0 RPM aligned to *S. italica*	19,298	20,248	19,893	20,041	19,814
Transcript number >1 RPM aligned to *S. italica*	14,875	15,378	16,001	15,910	15,654
Transcript number >20 RPM aligned to *S. italica*	4,882	5,263	5,826	5,807	5,495
Transcript number >1,000 RPM aligned to *S. italica*	132	143	135	136	136

Datasets

Dataset 1. Annotated transcriptome expression data (RPM) of *E. retroflexa* aquatic and terrestrial culms.

Dataset 2. CAP3 assembled filtered *E. retroflexa* unigene database.

Supplemental Material

Supplemental Datasets

Supplemental Dataset 1. Metabolic pathways covered by *E.retroflexa* unigene database.

Supplemental Data Set 2. Cellular processes covered by *E. retroflexa* unigene database.

Supplemental Dataset 3. Regulatory processes covered by *E. retroflexa* unigene database.

Supplemental Dataset 4. *E retroflexa* contigs annotated with Uniref100 via tblastx based on highest bitscore.

Supplemental Dataset 5. *E. retroflexa* contigs annotated against *S. italica* V2.1 primary transcripts.

Supplemental Dataset 6. Mapman category enrichment analysis (Fisher's Exact test).

Supplemental Dataset 7. Wilcoxon rank sum test of Mapman categories.

Supplemental Figures

Supplemental Figure 1. Histogram of *E. retroflexa* unigene database.

Supplemental Figure 2. Cumulative relative expression plots of reads mapped against various references by BLAT.

Supplemental Figure 3. Comparison of aquatic and terrestrial *E. retroflexa* transcriptomes.

Supplemental Figure 4. Photographic images of *E. retroflexa* cultivation.

Supplemental Figure 5. Quantitative transcript abundance patterns between aquatic and terrestrial *E. retroflexa* culms.

Supplemental Figure 6. Transcriptional investment of aquatic and terrestrial *E. retroflexa* culms.

Supplemental Figure 7. Transcriptional regulation and plant hormone expression patterns of aquatic and terrestrial *E. retroflexa* culms.

Supplemental Figure 8. Phylogeny of *E. retroflexa* based on internal transcribed spacer (ITS) sequences.

Supplemental Tables

Supplemental Table 1. Gene expression dynamics of *E. retroflexa* samples aligned to *S. italica* reference.

Supplemental Table 2. Overview of *E. retroflexa* CAP3 assembly statistics and annotation of contigs.

Supplemental Table 3. Pearson's correlation (*r*) of transcriptome data. *E. retroflexa* culms mapped to *S. italica*

Supplemental Table 4. Averaged C_4 cycle transcripts (RPMs) of *E. retroflexa* aquatic and terrestrial transcriptomes

Acknowledgements

Work in the authors' laboratory was supported by grants of the Deutsche Forschungsgemeinschaft (EXC 1028, IRTG 1525, and WE 2231/9-1 to APMW). We are grateful to the HHU Biomedical Research Center (BMFZ) for support with RNA-Seq analysis and to the MSU High Performance Computing Cluster (HPCC) for support with computational analysis of RNA-Seq data.

Author Contributions

C.K. performed experimental work, analyzed data and wrote the paper; S.S. set up growth conditions for plants and cultivated plants, took photographic images of plants and performed CAP3 assembly and bioinformatic analyses; M.S. performed analysis of transcriptome variability; A.K.D. performed relative cumulative expression and Edge R analyses; A.H. assisted with set up of growth conditions in aquaria; C.R.B assisted in data analysis; A.B. co-wrote the paper, assisted in data analysis and experimental design; A.P.M.W. designed study and co-wrote the paper.

References

Agarie, S., Kai, M., Takatsuji, H. and Ueno, O. (2002) Environmental and hormonal regulation of gene expression of C_4 photosynthetic enzymes in the amphibious sedge Eleocharis vivipara. *Plant Science* 163: 571-580.

Ashton, A., Burnell, J., Furbank, R., Jenkins, C. and Hatch, M. (1990) Enzymes of C_4 photosynthesis. In 'Methods in Plant Biochemistry, Volume 3,.(Ed. PJ Lea) pp. 39-72. Academic Press: San Diego, CA.

Bailey-Serres, J. and Voesenek, L.A. (2008) Flooding stress: acclimations and genetic diversity. *Annual Review of Plant Biology* 59: 313-339.

Bailey, S., Horton, P. and Walters, R.G. (2004) Acclimation of *Arabidopsis thaliana* to the light environment: the relationship between photosynthetic function and chloroplast composition. *Planta* 218: 793-802.

Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., et al. (2005) The universal protein resource (UniProt). *Nucleic Acids Research* 33: D154-D159.

Bennetzen, J.L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A.C., et al. (2012) Reference genome sequence of the model plant Setaria. *Nat Biotech* 30: 555-561.

Besnard, G., Muasya, A.M., Russier, F., Roalson, E.H., Salamin, N. and Christin, P.A. (2009) Phylogenomics of C_4 Photosynthesis in Sedges (Cyperaceae): Multiple Appearances and Genetic Convergence. *Molecular Biology and Evolution* 26: 1909-1919.

Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A., et al. (2010) Manipulation of FASTQ data with Galaxy. *Bioinformatics* 26: 1783-1785.

Boeger, M.R.T. and Poulson, M.E. (2003) Morphological adaptations and photosynthetic rates of amphibious *Veronica anagallis-aquatica L.* (Scrophulariaceae) under different flow regimes. *Aquatic Botany* 75: 123-135.

Bowes, G., Ogren, W.L. and Hageman, R.H. (1971) Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochemical and Biophysical Research Communications* 45: 716-722.

Bowes, G., Rao, S.K., Estavillo, G.M. and Reiskind, J.B. (2002) C_4 mechanisms in aquatic angiosperms: comparisons with terrestrial C_4 systems. *Functional Plant Biology* 29: 379-392.

Bräutigam, A., Kajala, K., Wullenweber, J., Sommer, M., Gagneul, D., Weber, K.L., et al. (2010) An mRNA blueprint for C_4 photosynthesis derived from comparative transcriptomics of closely related C_3 and C_4 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 C_4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C_4 species. *Journal of Experimental Botany* 65: 3579-3593.

Casati, P., Lara, M.V. and Andreo, C.S. (2000) Induction of a C_4-like mechanism of CO_2 fixation in *Egeria densa*, a submersed aquatic species. *Plant Physiology* 123: 1611-1621.

Centritto, M., Loreto, F. and Chartzoulakis, K. (2003) The use of low CO_2 to estimate diffusional and non-diffusional limitations of photosynthetic capacity of salt-stressed olive saplings. *Plant Cell and Environment* 26: 585-594.

Cernusak, L.A., Ubierna, N., Winter, K., Holtum, J.A., Marshall, J.D. and Farquhar, G.D. (2013) Environmental and physiological determinants of carbon isotope discrimination in terrestrial plants. *New Phytol* 200: 950-965.

Chen, T., Zhu, X.-G. and Lin, Y. (2014) Major alterations in transcript profiles between C_3-C_4 and C_4 photosynthesis of an amphibious species *Eleocharis baldwinii*. *Plant Molecular Biology* 86: 93-110.

Chen, T.Y., Ye, R.J., Fan, X.L., Li, X.H. and Lin, Y.J. (2012) Identification of C_4 photosynthesis metabolism and regulatory-associated genes in *Eleocharis vivipara* by SSH. *Photosynthesis Research* 112: 215-215.

Cheng, S., van den Bergh, E., Zeng, P., Zhong, X., Xu, J., Liu, X., et al. (2013) The *Tarenaya hassleriana* genome provides insight into reproductive trait and genome evolution of crucifers. *Plant Cell* 25: 2813-2830.

Chu, C., Dai, Z., Ku, M.S. and Edwards, G.E. (1990) Induction of Crassulacean Acid Metabolism in the facultative halophyte *Mesembryanthemum crystallinum* by Abscisic Acid. *Plant Physiology* 93: 1253-1260.

Colmer, T.D. and Voesenek, L.A.C.J. (2009) Flooding tolerance: suites of plant traits in variable environments. *Functional Plant Biology* 36: 665-681.

Coplen, T.B., Brand, W.A., Gehre, M., Groning, M., Meijer, H.A.J., Toman, B., et al. (2006) New guidelines for delta C-13 measurements. *Anal. Chem.* 78: 2439-2441.

De Vos, M., Van Oosten, V.R., Van Poecke, R.M., Van Pelt, J.A., Pozo, M.J., Mueller, M.J., et al. (2005) Signal signature and transcriptome changes of *Arabidopsis* during pathogen and insect attack. *Mol Plant Microbe Interact* 18: 923-937.

Edwards, G.E. and Ku, M.S. (1987) Biochemistry of C_3-C_4 intermediates. *New York: Academic Press* In M.D. Hatch & N.K. Boardmann (Eds.): pp.275-325.

Frost-Christensen, H., JØRgensen, L.B. and Floto, F. (2003) Species specificity of resistance to oxygen diffusion in thin cuticular membranes from amphibious plants. *Plant, Cell & Environment* 26: 561-569.

Frost-Christensen, H. and Sand-Jensen, K. (1995) Comparative kinetics of photosynthesis in floating and submerged *Potamogeton* leaves. *Aquatic Botany* 51: 121-134.

Furbank, R.T. and Hatch, M.D. (1987) Mechanism of C_4 photosynthesis: the size and composition of the inorganic carbon pool in bundle sheath cells. *Plant Physiology* 85: 958-964.

Ge, X., Chen, H., Wang, H., Shi, A. and Liu, K. (2014) De novo assembly and annotation of *Salvia splendens* transcriptome using the Illumina platform. *PLoS ONE* 9: e87693.

Gowik, U., Bräutigam, A., Weber, K.L., Weber, A.P.M. and Westhoff, P. (2011) Evolution of C_4 Photosynthesis in the Genus Flaveria: How Many and Which Genes Does It Take to Make C_4? *Plant Cell* 23: 2087-2105.

Haberlandt, G. (1904) Physiologische pflanzenanatomie. W. Engelmann.

Hatch, M.D. (1987) C_4 photosynthesis – a uniqe blend of modified biochemistry, anatomy and ultrastructure. *Biochimica Et Biophysica Acta* 895: 81-106.

Hatch, M.D. and Slack, C.R. (1970) Photosynthetic Co2-Fixation Pathways. *Ann Rev Plant Physio* 21: 141-&.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A.P., et al. (2013) Predicting C_4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell* 153: 1579-1588.

Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6: 65-70.

Holmes, M. and Klein, W. (1987) The light and temperature environments. *SPEC. PUBL. BR. ECOL. SOC. 1987.*

Huang, X.Q. and Madan, A. (1999) CAP3: A DNA sequence assembly program. *Genome Research* 9: 868-877.

Inda, L.A., Torrecilla, P., Catalán, P., and Ruiz-Zapata, T. (2008) Phylogeny of Cleome L. and its close relatives Podandrogyne Ducke and Polanisia Raf. (Cleomoideae, Cleomaceae) based on analysis of nuclear ITS sequences and morphology. *Plant Sys. Evol*. 274, 111-126.

Jackson, M.B. (2008) Ethylene-promoted elongation: an adaptation to submergence stress. *Ann. Bot.* 101:229–48

Jahnke, L.S., Eighmy, T.T. and Fagerberg, W.R. (1991) Studies of Elodea nuttalii grown under photorespirator conditions. Photosynthetic characteristics. *Plant Cell and Environment* 14: 147-156.

Keeley, J.E. (1998) C_4 photosynthetic modifications in the evolutionary transition from land to water in aquatic grasses. *Oecologia* 116: 85-97.

Keeley, J.E. (1999) Photosynthetic pathway diversity in a seasonal pool community. *Functional Ecology* 13: 106-118.

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

Kim, J.M., To, T.K., Nishioka, T. and Seki, M. (2010) Chromatin regulation functions in plant abiotic stress responses. *Plant, Cell & Environment* 33: 604-611.

Kirk, J.T.O. (1994) Light and photosynthesis in aquatic ecosystems. Cambridge university press.

Krause-Jensen, D. and Sand-Jensen, K. (1998) Light attenuation and photosynthesis of aquatic plant communities. *Limnology and Oceanography* 43: 396-407.

Külahoglu, C., Denton, A.K., Sommer, M., Mass, J., Schliesky, S., Wrobel, T.J., et al. (2014) Comparative transcriptome atlases reveal altered gene expression modules between two Cleomaceae C_3 and C_4 plant species. *The Plant Cell* 26: 3243-3260.

Kuwabara, A., Tsukaya, H. and Nagata, T. (2001) Identification of factors that cause heterophylly in Ludwigia arcuata Walt. (Onagraceae). *Plant Biology* 3: 670-670.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* 40: D1202-1210.

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

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., et al. (2007) Clustal W and clustal X version 2.0. *Bioinformatics* 23: 2947-2948.

Lee, S.C., Mustroph, A., Sasidharan, R., Vashisht, D., Pedersen, O., Oosumi, T., et al. (2011) Molecular characterization of the submergence response of the *Arabidopsis thaliana* ecotype Columbia. *New Phytol* 190: 457-471.

Lee, W.Y., Lee, D., Chung, W.-I. and Kwon, C.S. (2009) Arabidopsis ING and Alfin1-like protein families localize to the nucleus and bind to H3K4me3/2 via plant homeodomain fingers. *Plant Journal* 58: 511-524.

Long, S.P. and Bernacchi, C.J. (2003) Gas exchange measurements, what can they tell us about the underlying limitations to photosynthesis? Procedures and sources of error. *Journal of Experimental Botany* 54: 2393-2401.

Maberly, S.C. and Madsen, T.V. (2002) Freshwater angiosperm carbon concentrating mechanisms: processes and patterns. *Functional Plant Biology* 29: 393-405.

Maberly, S.C. and Spence, D.H.N. (1989) Photosynthesis and photorespiration in fresh-water organisms – amphibious plants. *Aquatic Botany* 34: 267-286.

Madsen, T.V. and Sandjensen, K. (1991) Photosynthetic carbon assimilation in aquatic macrophytes. *Aquatic Botany* 41: 5-40.

Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M.J., Weber, A.P., Westhoff, P., et al. (2014) The role of photorespiration during the evolution of C_4 photosynthesis in the genus Flaveria. *Elife*: e02478.

Martin, J.A. and Wang, Z. (2011) Next-generation transcriptome assembly. *Nature Reviews Genetics* 12: 671-682.

McElwain, E.F., Bohnert, H.J. and Thomas, J.C. (1992) Light moderates the induction of phosphoenolpyruvate carboxylase by NaCl and abscisic acid in Mesembryanthemum crystallinum. *Plant Physiology* 99: 1261-1264.

Micheli, F. (2001) Pectin methylesterases: cell wall enzymes with important roles in plant physiology. *Trends in Plant Science* 6: 414-419.

Minorsky, P.V. (2003) The hot and the classic. *Plant Physiology* 133: 1671-1672.

Mommer, L., de Kroon, H., Pierik, R., Bogemann, G.M. and Visser, E.J.W. (2005a) A functional comparison of acclimation to shade and submergence in two terrestrial plant species. *New Phytologist* 167: 197-206.

Mommer, L., Pons, T.L. and Visser, E.J. (2006) Photosynthetic consequences of phenotypic plasticity in response to submergence: Rumex palustris as a case study. *Journal of Experimental Botany* 57: 283-290.

Mommer, L., Pons, T.L., Wolters-Arts, M., Venema, J.H. and Visser, E.J. (2005b) Submergence-induced morphological, anatomical, and biochemical responses in a terrestrial species affect gas diffusion resistance and photosynthetic performance. *Plant Physiology* 139: 497-508.

Mommer, L. and Visser, E.J. (2005) Underwater photosynthesis in flooded terrestrial plants: a matter of leaf plasticity. *Annals of Botany* 96: 581-589.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C. and Kanehisa, M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* 35: W182-W185.

Mustroph, A., Zanetti, M.E., Jang, C.J., Holtan, H.E., Repetti, P.P., Galbraith, D.W., et al. (2009) Profiling translatomes of discrete cell populations resolves altered cellular priorities during hypoxia in Arabidopsis. *Proc Natl Acad Sci U S A* 106: 18843-18848.

Papanicolaou, A., Stierli, R., Ffrench-Constant, R.H. and Heckel, D.G. (2009) Next generation transcriptomes for next generation genomes using est2assembly. *BMC Bioinformatics* 10: 447.

Pedersen, O., Colmer, T.D. and Sand-Jensen, K. (2013) Underwater photosynthesis of submerged plants - recent advances and methods. *Front Plant Sci* 4: 140.

Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* 29.

Pigliucci, M. (2001) Phenotypic plasticity: beyond nature and nurture. *Syntheses in ecology and evolution*.

Porra, R.J., Thompson, W.A. and Kriedemann, P.E. (1989) Determination of accurate extinction coefficients and simultaneous-equations for asssaysing chlorophyll a and chlorophyll b extracted with 4 different solvents – verification of the concentration of chlorophyll standards by atomic-absorption spectroscopy. *Biochimica Et Biophysica Acta* 975: 384-394.

R Development Core Team (2009). R: A Language and Environment for Statistical Computing. (Vienna, Austria: R Foundation for Statistical Computing).

Reiskind, J.B., Madsen, T.V., Van Ginkel, L.C. and Bowes, G. (1997) Evidence that inducible C_4-type photosynthesis is a chloroplastic CO_2-concentrating mechanism in Hydrilla, a submersed monocot. *Plant, Cell & Environment* 20: 211-220.

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

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

Saeed, A.I., Hagabati, N.K., Braisted, J.C., Liang, W., Sharov, V., Howe, E.A., et al. (2006) TM4 microarray software suite. In *DNA Microarrays, Part B: Databases and Statistics*. Edited by Kimmel, A. and Oluver, B. pp. 134-93.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., et al. (2003) TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* 34: 374-8.

Sage, R.F. (2004) The evolution of C_4 photosynthesis. *New Phytologist* 161: 341-370.

Sage, R.F. and McKown, A.D. (2006) Is C_4 photosynthesis less phenotypically plastic than C_3 photosynthesis? *Journal of Experimental Botany* 57: 303-317.

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

Sand-Jensen, K. (1989) Environmental variables and their effect on photosynthesis of aquatic plant communities. *Aquatic Botany* 34: 5-25.

Sand-Jensen, K. and Frost-Christensen, H. (1999) Plant growth and photosynthesis in the transition zone between land and stream. *Aquatic Botany* 63: 23-35.

Schliesky, S., Gowik, U., Weber, A.P. and Brautigam, A. (2012) RNA-Seq Assembly - Are We There Yet? *Front Plant Sci* 3: 220.

Schmittgen, T.D. and Livak, K.J. (2008) Analyzing real-time PCR data by the comparative C-T method. *Nat Protoc* 3: 1101-1108.

Schulze, S., Mallmann, J., Burscheidt, J., Koczor, M., Streubel, M., Bauwe, H., et al. (2013) Evolution of C_4 Photosynthesis in the Genus *Flaveria*: Establishment of a Photorespiratory CO_2 Pump. *Plant Cell* 25: 2522-2535.

Sculthorpe, C. (1967) The biology of vascular plants. *Edward Arnold, London* 610.

Setter, T.L. and Laureles, E.V. (1996) The beneficial effect of reduced elongation growth on submergence tolerance of rice. *Journal of Experimental Botany* 47: 1551-1559.

Sommer, M., Brautigam, A. and Weber, A.P. (2012) The dicotyledonous NAD malic enzyme C_4 plant *Cleome gynandra* displays age-dependent plasticity of C_4 decarboxylation biochemistry. *Plant Biology* 14: 621-629.

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

Tsuji, H., Saika, H., Tsutsumi, N., Hirai, A. and Nakazono, M. (2006) Dynamic and reversible changes in histone H3-Lys4 methylation and H3 acetylation occurring at submergence-inducible genes in rice. *Plant and Cell Physiology* 47: 995-1003.

Ueno, O. (2001) Environmental Regulation of C_3 and C_4 Differentiation in the Amphibious Sedge *Eleocharis vivipara*. *Plant Physiology* 127: 1524-1532.

Ueno, O. (2004) Environmental regulation of photosynthetic metabolism in the amphibious sedge *Eleocharis baldwinii* and comparisons with related species. *Plant Cell and Environment* 27: 627-639.

Ueno, O., Samejima, M. and Koyama, T. (1989) Distribution and evolution of C_4 syndrome in Eleocharis, a sedge group inhabiting wet and aquatic environments, based on culm anatomy and carbon isotope ratios. *Annals of Botany* 64: 425-438.

Ueno, O., Samejima, M., Muto, S. and Miyachi, S. (1988) Photosynthetic characteristics of an amphibious plant, *Eleocharis vivipara*-Expression of C_4 and C_3 modes in contrasting environments. *Proc Natl Acad Sci U S A* 85: 6733-6737.

Ueno, O. and Wakayama, M. (2004) Cellular expression of C_3 and C_4 photosynthetic enzymes in the amphibious sedge Eleocharis retroflexa ssp. chaetaria. *J Plant Res* 117: 433-441.

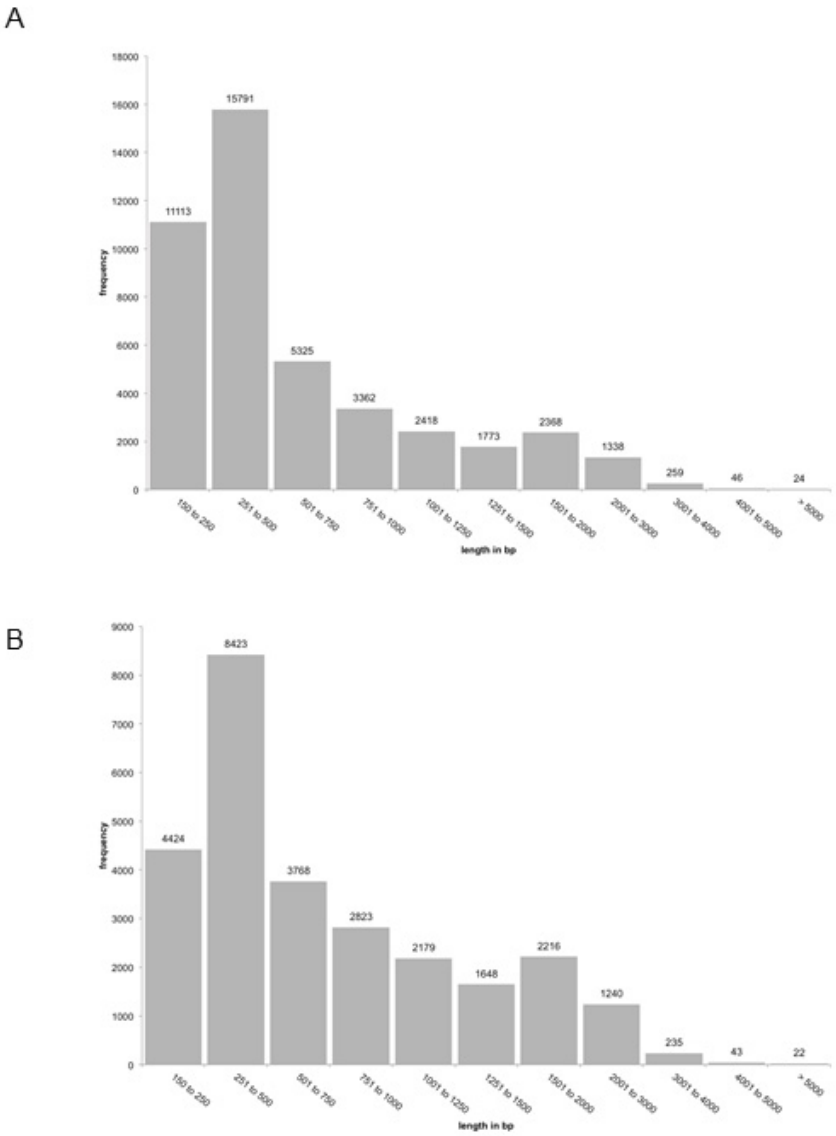
Usadel, B., Nagel, A., Thimm, O., Redestig, H., Blaesing, O.E., Palacios-Rojas, N., et al. (2005) Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of coresponding genes, and comparison with known responses. *Plant Physiology* 138: 1195-1204.

Vervuren, P.J.A., Blom, C.W.P.M. and De Kroon, H. (2003) Extreme flooding events on the Rhine and the survival and distribution of riparian plant species. *Journal of Ecology* 91: 135-146.

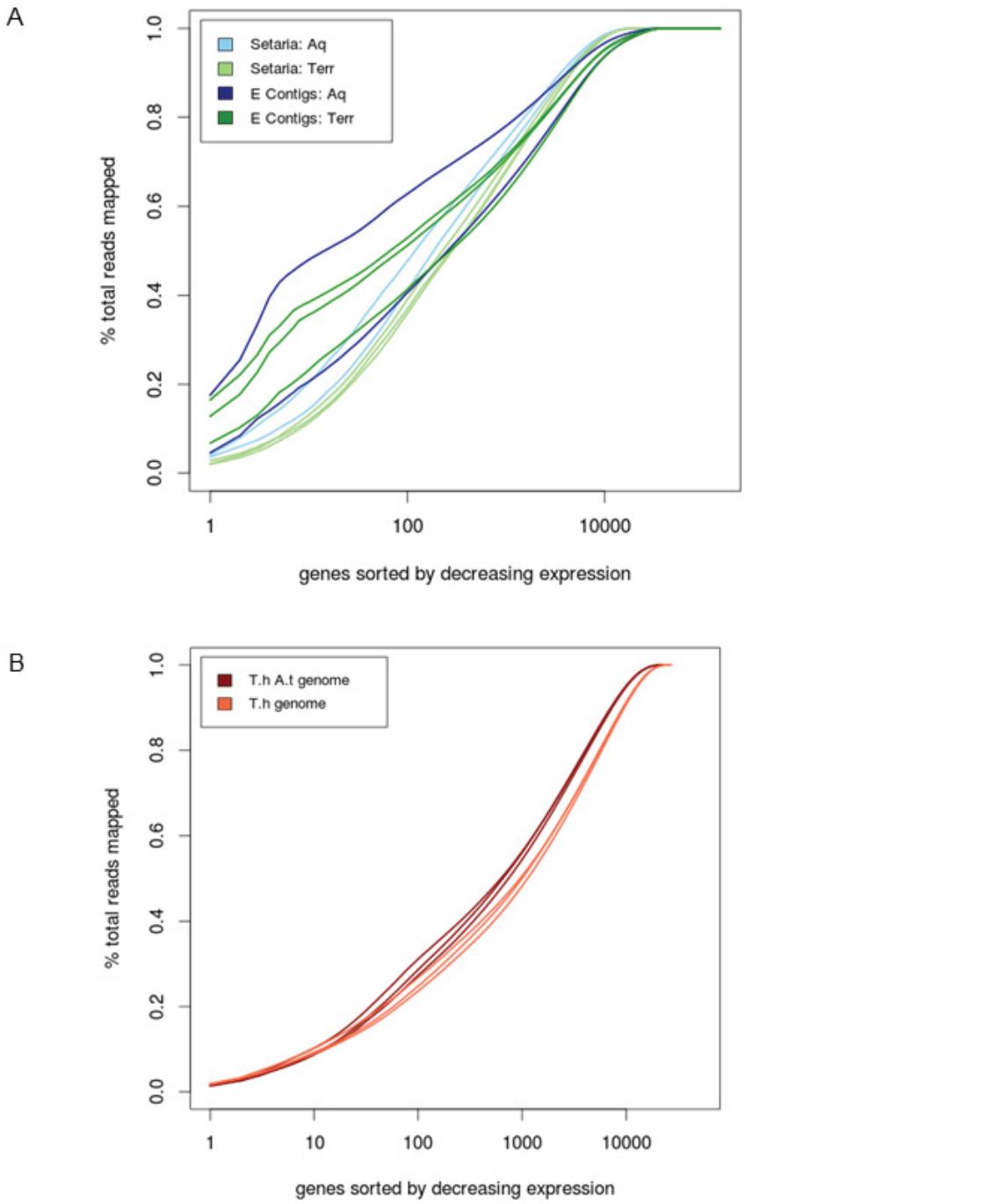
Wang, P., Kelly, S., Fouracre, J.P. and Langdale, J.A. (2013) Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C_4 Kranz anatomy. *Plant Journal* 75: 656-670.

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

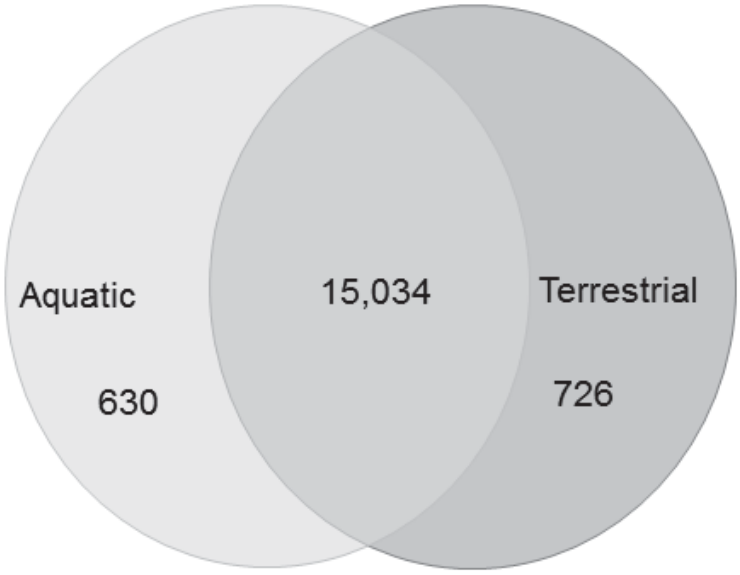
Yekutieli, D. and Benjamini, Y. (1999) Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference* 82: 171-196.



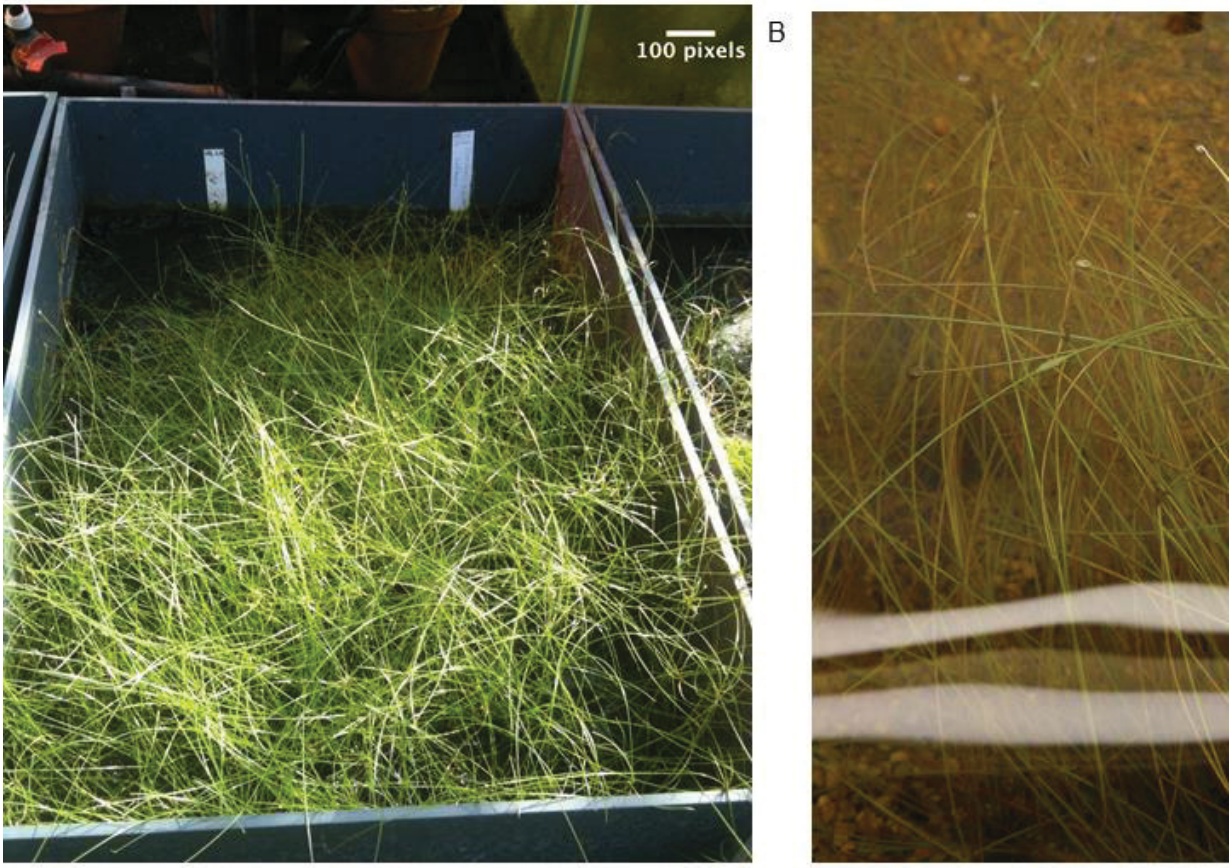
Supplemental Figure 1. Histogram of *E. retroflexa* unigene database.
(A) Frequency distribution of *E. retroflexa* contigs (43,817) with length over 200 bp assembled with CAP3. **(B)** Frequency distribution of filtered *E. retroflexa* contig length of contigs (27,021), that were annotated by either *S. italica* or UniRef100 as *Viridiplantae*.



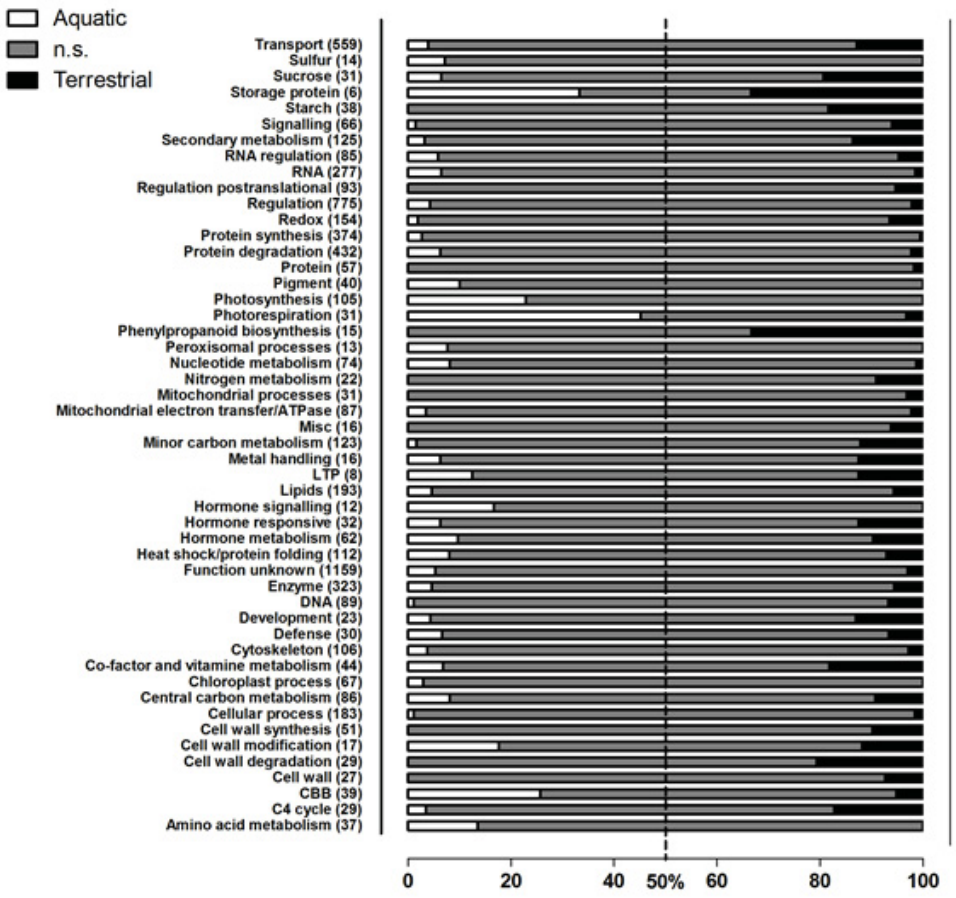
Supplemental Figure 2. Cumulative relative expression plots of reads mapped against various references by BLAT. Raw count files were sorted descending by matched reads per contig/cds and the amount of reads relative to total mapped reads per sample was summed up. **(A)** Submerse aquatic (blue) and terrestrial (green) *E. retroflexa* reads mapped to *E. retroflexa* contigs (darker color) or *S. italica* primary transcript (lighter color). **(B)** *Tarenaya hassleriana* mature leaf reads mapped to the *T. hassleriana* primary transcripts (lighter color; T.h) and the *Arabidopsis thaliana* reference (dark color; A.t).



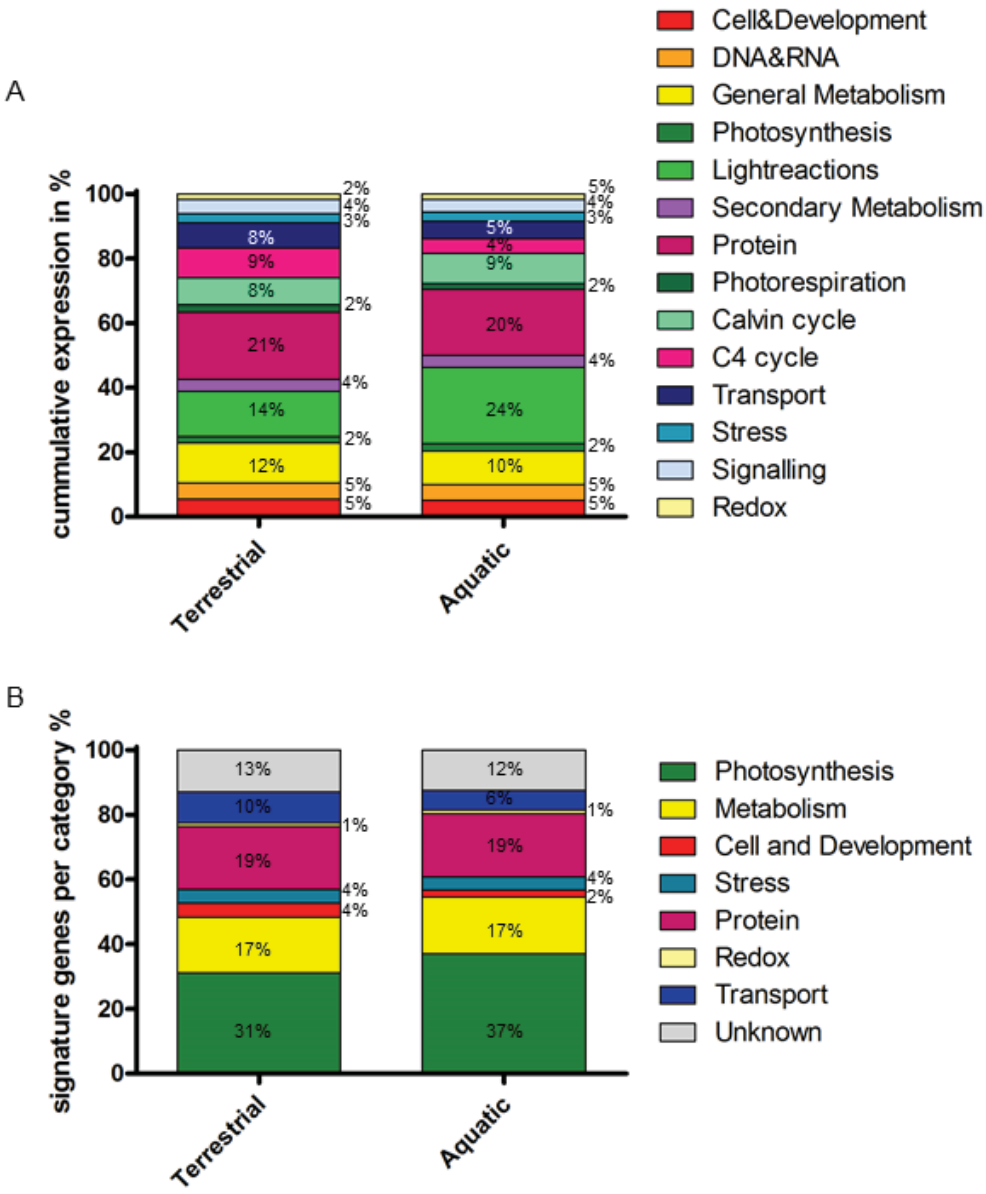
Supplemental Figure 3. Comparison submerse aquatic and terrestrial *E. retroflexa* transcriptomes. Numbers indicate transcripts that are the same between averaged *E. retroflexa* transcriptomes (n=3, terrestrial, green; n=2, aquatic, blue) or significantly higher abundant 8% of all transcripts (P-value<0.05) in either the aquatic and the terrestrial sample (92%). *P*-values were Benjamini-Hochberg FDR corrected.



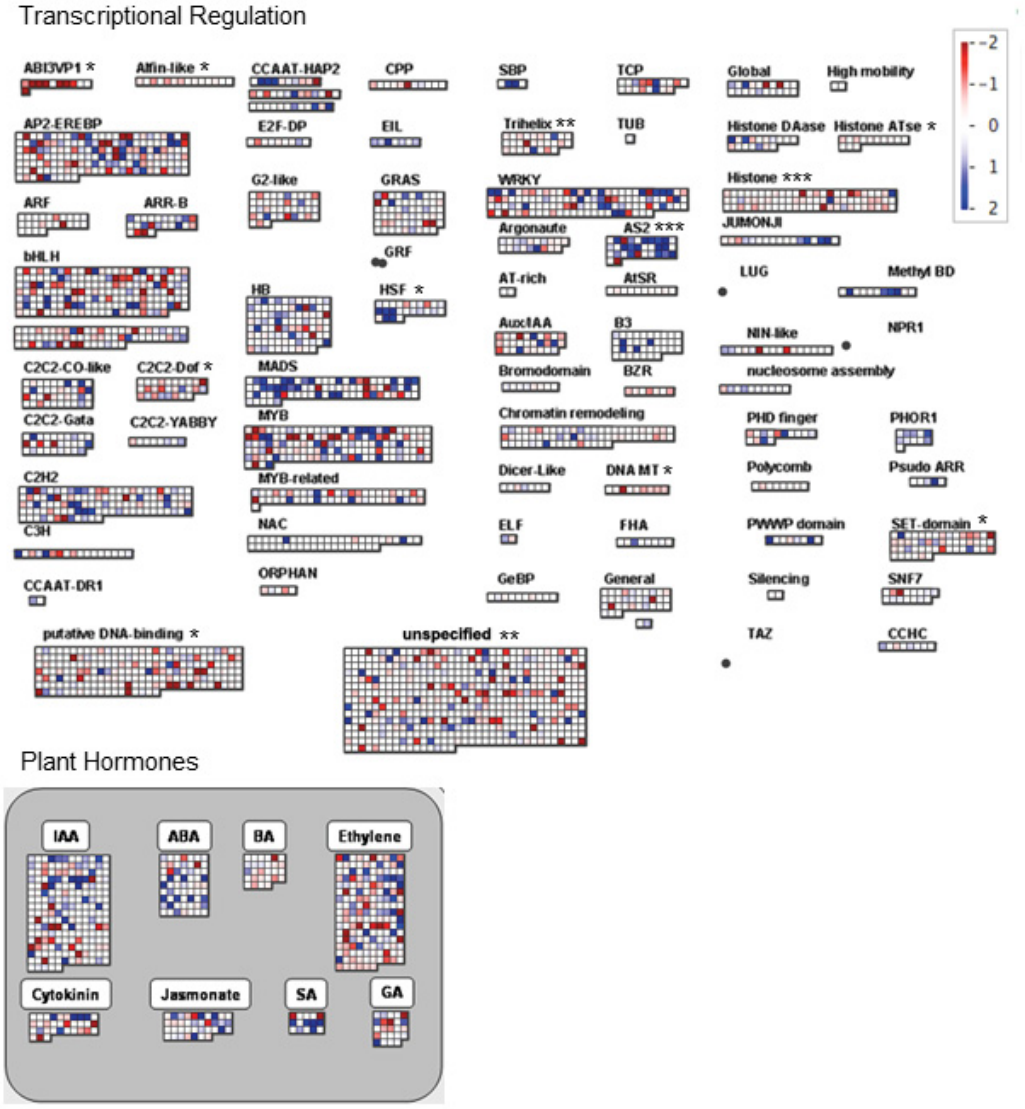
Supplemental Figure 4. Photographic images of *E. retroflexa* cultivation. (A) Phenotype of 6 week old terrestrial *E. retroflexa* grown in swamp boxes**. (B)** Phenotype of submerged E. *retroflexa* plants grown in aquaria after 4 weeks of cultivation. Scale bar represents 100 pixel.



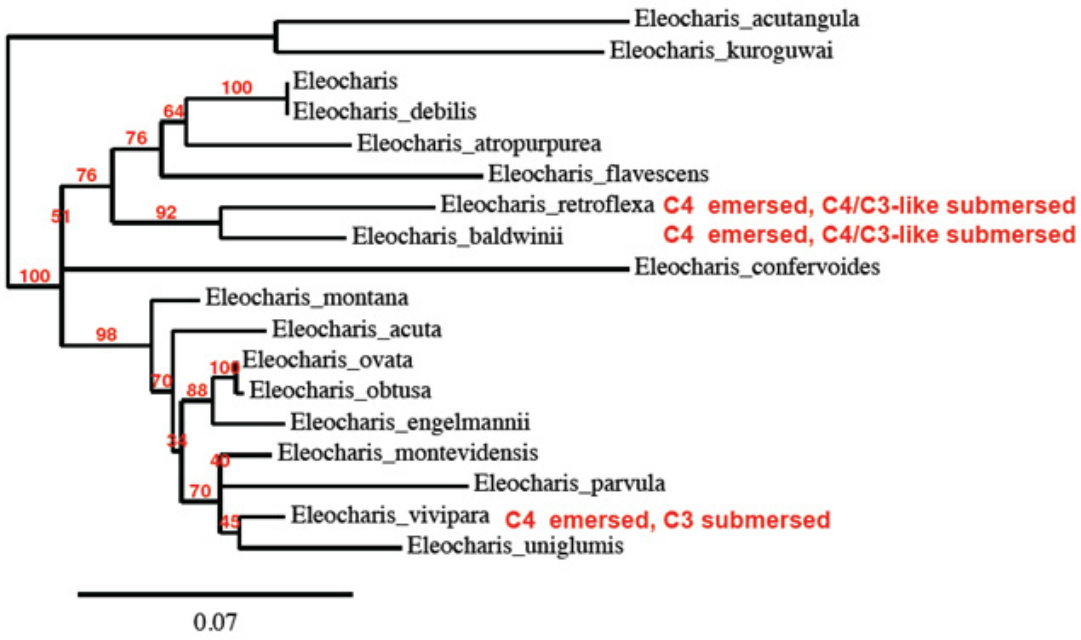
Supplemental Figure 5. Quantitative transcript abundance patterns between aquatic and terrestrial *E. retroflexa* culms. Relative number of transcripts (in percent), that are significantly up-regulated (BH-corrected P-value<0.05) in either submersed aquatic (white) or terrestrial (black) transcriptome per custom mapman-derived category. Numbers of all genes per category are indicated in parentheses.



Supplemental Figure 6. Transcriptional investment of *E. retroflexa* culms grown under aquatic and terrestrial conditions.
(A) Cumulative average RPMs in percent of custom basal Mapman categories for each tissue in *E. retroflexa*. **(B)** Distribution signature genes in *E. retroflexa* terrestrial and submerse aquatic culms. Percentage of signature genes expressed over 1,000 RPM falling in each basal Mapman category for every averaged tissue.



Supplemental Figure 7.Transcriptional regulation and plant hormone expression patterns in *E. retroflexa* culms grown under water and on soil.
Overview of gene expression dynamics in the Mapman categories (TAIR10) plant hormones and transcriptional regulation in *E. retroflexa* culms. Heatmaps depict transcriptional log2 fold-changes of submerse aquatic versus terrestrial transcriptomes (RPM). Red (ratio<0) represents an increase of gene expression in submerse aquatic and blue (ratio>0) an increase of transcript accumulation in terrestrial culms. Significant fold-changes were calculated by Wilcoxon Rank Test and are indicated by asterisk. P-values (* P-value < 0.05; ** P-value < 0.01; ***P-value < 0.001) were Benjamini-Yekutieli FDR corrected.



Supplemental Figure 8. Phylogeny of *E. retroflexa* based on internal transcribed spacer (ITS) sequences.
Phylogenetic tree is based on ITS sequence similarity. Tree is based on Maximum likelihood and alignments. Red numbers indicate branch support by 100 times bootstrapping.

Ich bedanke mich herzlich bei:

Prof. Dr. Andreas P. M. Weber für die Chance, die Projektarbeit während meiner Dissertation mit großer kreativer Freiheit durchführen zu dürfen, und dafür, sie an den nötigen Stellen zu hinterfragen und zu lenken. Vielen Dank, dass Sie mich in Ihrem Labor aufgenommen haben und ich hier die Grundlagen wissenschaftlichen Arbeitens und zahlreiche Methoden auch schon während meiner Diplomarbeit erlernen durfte. Sie haben mir während dieser Zeit im Rahmen meiner Beschäftigung alle Möglichkeiten gegeben, mir viele Türen geöffnet und außerdem meine wissenschaftlichen Ideen außerhalb des Promotionsvorhabens gefördert. Danke dafür.

Prof. Dr. Peter Westhoff für zahlreiche Begutachtungen und Beurteilungen meiner Arbeit einschließlich meiner Dissertation und für das große Vertrauen in meine Person und Arbeit, die diese Gutachten mit sich bringen.

Dr. Andrea Bräutigam dafür, dass du sowohl die Rolle der Betreuerin meiner Diplomarbeit eingenommen hast, als auch die eines interessierten wissenschaftlichen Gesprächspartners, den ich besonders im späteren Verlauf meiner Dissertation gebraucht und gerne konsultiert habe. Deine sachliche, wissenschaftliche Art zu diskutieren hat meine Projekte bereichert und letztendlich mit zu dem gemacht was sie jetzt sind. Ich danke dir außerdem für das Korrekturlesen meiner Manuskripte.

Dr. Rosalia Deeken und Stefanie Saupe für eure Gastfreundschaft in Würzburg und dafür, dass ich die „Fat-Western“ Methode von euch erlernen durfte.

Prof. Dr. Jürgen Zeier und Elia Stahl für die nette Hilfe bei der Messung von Phytosterolen in Pflanzenextrakten.

Tobias Schumann für die Zusammenarbeit bei den Quenching-Analysen von *glk* Mutanten.

Den Mass-Spec Ladies Dr. Tabea Mettler-Altmann, Katrin Weber, Elisabth Klemp und Maria Graf für die Metabolit-Analysen und eure Hilfsbereitschaft bei der Auswertung. Danke Katrin für das Lab-Mentoring zu meiner Anfangszeit und danke Tabbi für informative Gespräche über Massenspektrometer und Gott und die Welt.

Iniyan Ganesan und James Rauschendorfer, dass ihr eure Sommerferien in den USA damit verbracht habt, mich bei der Arbeit an meinen Projekten zu unterstützen.

Lisa Leson, dass du mich als Neuling unter deine Fittiche genommen und angelernt hat, woraus sich eine dicke Freundschaft entwickelt hat.

Dominik Brilhaus für deine Hilfe bei der Auswertung meiner RNA-Sequenzanalysen, aber eigentlich viel mehr für 31543267 Kaffee-, Tee-, Mittags-, Skat-, Nachmittags und Frühstückspausen, 432234 Abende mit Gesellschaftsspielen, Pommes in nassen Jacken, Lebensweisheiten am Rheinufer austauschen oder auch mal einem Bierchen und Zigarillo und für alle anderen gemeinsamen Erlebnisse, die Düsseldorf für mich zu einem schöneren Flecken Erde gemacht haben.

Pia Dahlhoff für deine verrückte Art und deine große Hilfsbereitschaft.

Der C_4-Gruppe für anregende Diskussionen über den besten Photosynthesetyp der Welt bei mehr oder weniger wöchentlichen Meetings.

Dem Weberlab für seine persönliche und herzliche Art und alle seine Mitglieder, die das ermöglichen und immer ein offenes Ohr für Fragen haben.

Allen, die an der Erstellung der Manuskripte beteiligt waren und die ich nicht namentlich nicht einzeln erwähnen konnte.

Dr. Thea Pick. Du vervollständigst mich.

Meiner Familie für die tollste Unterstützung, die ich mir während meiner Ausbildung wünschen konnte.

Dem DFG Priority Programme „Adaptomics“ für die Finanzierung.